Rapport remis à l’OPESCT

Calcul d’indicateurs statistiques sur les effets indésirables liés aux injections ARNm depuis les bases de données de pharmacovigilance européenne (EudraVigilance) et américaine (VAERS)

29 avril 2022

Vincent Pavan, Enseignant-chercheur en mathématique, Université d’Aix-Marseille

Emmanuelle Darles, Enseignante-chercheuse informatique, Université de Poitiers

Table des matières

Résumé

Ce rapport présente les résultats obtenus sur l'impact de la vaccination avec des injections de type ARNm Pfizer BioNTech et Moderna depuis les données européennes (EudraVigilance) et américaines (VAERS). Les méthodes proposées permettent d'établir une balance bénéfice / risque quantitative selon le sexe et la classe d'âge mais également de fournir des signaux d'alerte forts notamment sur la fréquence d'apparition de certaines pathologies lourdes pouvant être imputables à la vaccination (cécité, choc anaphylactique, AVC) mais également fréquemment observées en population générale telle que les troubles du cycle menstruel. Les détails mathématiques des méthodes utilisées sont fournis en annexe, ainsi que le code d'analyse et de traitement informatique, garantissant transparence et reproductibilité.

Table des figures

Liste des tableaux

1 Introduction

1.1 l'imputabilité au sens du Professeur Bégaud

C'est par une citation du Ministre de la santé, Monsieur Olivier Véran, le 15 octbore 2021 devant la commission des lois de l'Assemblée Nationale [7] qu'il convient d'introduire notre propos :

Il n'y a pas, à ce stade, de cas de décès avéré imputable à la vaccination

Il y a alors deux remarques que l'on peut formuler :

1. Cette affirmation est parfaitement vraie;
2. Elle illustre exactement les faiblesses de la pharmacovigilance française sur la question vaccinale.

En effet, le terme d'imputabilité (intrinsèque) possède une signification très précise en pharmacovigilance française : il s'agit d'un schéma dit "challenge - déchallenge - rechallenge " que l'on peut résumer comme suit :

1. On prend un médicament et un effet indésirable apparaît : c'est l'étape dite de challenge;
2. On arrête la prise du médicament et l'effet indésirable disparaît : c'est l'étape dite de déchallenge;
3. On reprend le médicament et l'effet indésirable réapparaît : c'est l'étape de rechallenge.

Chacune des étapes est considérée comme une variable (C, D, R) qui peut prendre plusieurs modalités (ou valeurs). Ainsi dans l'étape de déchallenge, on distingue l'évolution suggestive, non concluante, non suggestive. Finalement on construit, à partir des variables (C,D,R) le score d'imputabilité chronologique :

$$\mathscr{C} = f\left(C, D, R\right), \quad \textit{où } \mathscr{C} \textit{ est le score d'imputabilité chronologique} \tag{1}$$

Il y a 4 modalités possible au score d'imputabilité chronologique que l'on note respectivement $(\mathscr{C}_i, i :\in [0,3])$

Ces critères d'imputabilité chronologique sont complétés par des considérations dites "sémiologiques", qui permettent d'évaluer la nature de l'effet indésirable et la manière avec lequel il a été diagnostiqué. Quatre modalités sont retenues pour les critères sémiologiques concernant la nature de l'effet indésirable. Il s'agit respectivement du mécanisme d'action M, du facteur favorisant F, du diagnostic différentiel D et des examens complémentaires E. A partir des valeurs associées à ces modalités (variables), on construit une fonction permettant d'évaluer un score de sémiologie :

$$\mathscr{S} = g\left(M, F, D, E\right), \quad \textit{où } \mathscr{S} \textit{ est le score d'imputabilité sémiologique} \tag{2}$$

Il y a 3 valeurs possibles au score d'imputabilité sémiologique que l'on note respectivement $(\mathscr{S}_i, i :\in [1,3])$

Finalement, en fonction des valeurs $\left(\mathscr{C}_i, \mathscr{S}_j\right)$ associées à la chronologie et à la sémiologie, on définit un score d'imputabilité $(\mathscr{I}_k : k \in [0,4])$ qui permet de qualifier une forme de causalité. Ces inputabilités s'échelonnent de "très vraisemblable" à "incompatible".

On pourra se reporter à https://pharmacovigilance-npdc.fr/enseignement-formation-pharmacologie/imputabilite-medicamenteuse-begaud/ pour un descriptif synthétique de la méthode.

Hélas cette notion d'imputabilité n'a jamais aucune implication dans l'étude de la vaccination car l'étape de déchallenge ou de rechallenge ne peuvent jamais survenir (on ne peut pas arrêter, par principe, d'avoir pris un vaccin). La conséquence, c'est que dans la tabulation même de la méthode, l'effet indésirable survenu au cours du challenge aura toujours une imputabilité douteuse ou incompatible avec la prise de l'injection.

C'est essentiellement cette considération qui permet au Ministre de la Santé Olivier Véran d'annoncer que les décès post-vaccinaux ne sont pas imputables aux vaccins.

Un des problèmes majeurs de cette "méthode" de pharmacovigilance (par ailleurs aussi appelée la méthode Bégaud) c'est qu'elle s'applique de façon obligatoire à l'ensemble des CRPV (***C****entre* ***R****égionaux de* ***P****harmaco****V****igilance*) en France. Ainsi sur https://pharmacovigilance-npdc.fr/enseignement-formation-pharmacologie/imputabilite-medicamenteuse-begaud/, le CRPV de la région Nord-Pas de Calais rappelle que :

La méthode Bégaud est obligatoire pour les CRPV en France

Accueil | Brèves en pharmacovigilance | Histoire de la pharmacovigilance | Formation | Manifestations, publications | Liens | Autres vigilances

Centre Régional de Pharmacovigilance
Nord-Pas de Calais

Que faut-il déclarer ? | Déclarer un effet indésirable (Professionnel) | Déclarer un effet indésirable (Patient) | Devenir de votre déclaration

Méthode française d'imputabilité médicamenteuse, dite méthode Bégaud

Michael Comments 2 juillet 2014 Méthode française d'imputabilité médicamenteuse, dite méthode Bégaud - Dernière modification : 2017-05-11T22:20:24+02:00 5

La méthode Bégaud est obligatoire pour les CRPV en France. Elle a été élaborée en 1978 par Dangoumo et al. et **actualisée en 1985 par Bégaud et al**. Elle distingue l'imputabilité intrinsèque et l'imputabilité extrinsèque.

L'imputabilité intrinsèque établit la relation de cause à effet entre chaque médicament et la survenue d'un effet indésirable, sur 3 critères **chronologiques** (tableaux 1 et 2) et 4 critères **sémiologiques** (tableaux 3 et 4). Il existe des abaques permettant d'associer ces critères en un score d'imputabilité intrinsèque (tableau 5). **L'imputabilité extrinsèque** recherche les cas similaires dans la littérature. Elle repose sur des critères bibliographiques (tableau 6). L'ensemble des critères est résumé dans le tableau 7.

Contact

Centre Régional de Pharmacovigilance (CRPV)
Nord Pas de Calais
CHRU de Lille
1, place de Verdun
59045 Lille Cedex
Tél : 03 20 44 68 64
Fax : 03 20 44 56 87
E-mail : pharmacovigilance@chru-lille.fr

Si cette affirmation est vraie, elle est inquiétante, s'agissant en particulier de comprendre et d'identifier les signaux liés aux dangers potentiels de certains vaccins.

1.2 Faiblesse formelle de la méthode Bégaud

La notion d'imputabilité est une notion de causalité dans un environnement incertain. Elle est d'ordre tautologique (l'imputabilité n'a aucun sens en dehors de sa définition) et donc ne peut avoir d'interprétation intuitive autre que celle d'une forme de causalité, mais en évitant précisément de s'appeler causalité (ce qui brouille la compréhension et permet mal d'orienter la décision politique). Il s'agit d'une notion purement qualitative, qui ne doit rien à la théorie des statistiques ou des probabilités. *Il n'y a pas de notion chiffrée autour de cette méthode.* Elle n'appartient pas au domaine quantitatif et évolue dans le flou et dans l'incertain. De cette façon, elle est intrinsèquement - selon l'expression - à géométrie variable. Son application relève de l'expérience, c'est à dire de l'habitude et des règles informelles que l'on a ou pas de pratiquer la méthode.

N'appartenant pas au domaine du calcul numérique, ni à celui du calcul formel (logique ou symbolique) la méthode d'imputabilité de Bégaud possède assez peu de fiabilité et de solidité scientifique. Il s'agit en réalité d'une pratique conventionnelle, dont la source de légitimité doit être comprise du point de vue de la hiérarchie universitaire et de la position d'autorité bien plus que du côté de la science.

Un document en annexe (voir annexe .1) montre que d'un point de vue strictement logique, la méthode est

intrinsèquement contradictoire (l'interprétation intuitive en est la suivante : si on a pris un médicament, on ne peut plus dire ensuite qu'on ne l'a pas pris). De fait un modèle en logique formelle (logique des propositions) montre sans ambiguïté que la méthode d'imputabilité, lorsqu'elle suit le challenge - déchallenge, permet de conclure l'imputabilité tout aussi bien que l'absence d'imputablité !

De façon générale, il apparaît que la culture de la pharmacovigilance en France - depuis l'apparition de la méthode challenge/déchallenge/rechallenge - ne retient aucune notion de formalisation numérique ou logique. Dans un échange mail que nous avons pu avoir avec Bernard Bégaud lui-même (voir annexe .1), nous avons mis en lumière l'incapacité de cette personne à formuler un simple problème de biais d'échantillonnage de façon mathématisée. On ne sera donc pas surpris, que dans une thèse supervisée par Bernard Bégaud - intitulée pourtant *Les modèles quantitatifs pour l'évaluation du rapport bénéfice-risque des vaccins* [2] - on ne trouve en réalité aucune formule mathématique, et que l'on se contente de la définition de quelques ratios à partir de tableaux de contingence. Cette thèse n'évite pas les poncifs sur l'immunité collective, dont nous avons déjà dénoncé l'imposture mathématique et scientifique.

La thèse que nous avons citée résume également la position de Bernard Bégaud et de la pharmacovigilance française en général sur l'existence de base de données, qui permettent la remontée des effets indésirables de toute nature suite à des vaccinations ([2], p 41).

Pour être analysées, les notifications spontanées sont regroupées dans des bases de données nationales (ex : base de données de la pharmacovigilance française gérée par l'ANSM), ou internationales (telles que Eudravigilance ou la base de l'OMS, WHO-UMC, à Uppsala en Suède). Dans certains pays, il existe un système de recueil des notifications de pharmacovigilance spécifique pour les vaccins. On peut citer par exemple le « Vaccine Adverse Event Reporting System » (VAERS) aux États-Unis. La principale qualité de la surveillance passive, malgré l'existence d'une sous-notification généralement forte, est sa sensibilité pour détecter un signal. Il est cependant impossible de quantifier valablement un risque à partir des données recueillies (existence d'une sous-notification généralement importante, absence de données précises sur la population exposée, biais de notification sélective, qualité de l'information reportée très variable). Ainsi, la surveillance passive doit être considérée comme un outil de détection de signaux et de génération d'hypothèses devant être complétées par d'autres études utilisant différentes méthodologies afin de valider, ou non, l'existence d'une association entre un vaccin et un événement indésirable.

Comme on le voit, la pharmacovigilance française accorde peu d'intérêt aux bases de données rapportant les évènements indésirables recueillis de façon spontanée. Il semble que le principal reproche que la pharmacovigilance classique puisse faire à ces remontées, c'est qu'elles sont "biaisées". Ainsi, dans l'échange que nous avons pu avoir avec ce dernier, et alors qu'il déniait contre toute forme de rigueur scientifique, notre travail sur les données du VAERS et d'EUDRAVIGILANCE, Monsieur Bégaud précisait-il que :

Cependant, l'analyse VAERS que vous publiez est un non sens méthodologique (volontaire ou non) absolu. Ce que vous dites serait vrai (c'est du reste l'une de nos méthodes de détection des signaux) SI et SEULEMENT SI on était dans le cadre d'une surveillance épidémiologique dans laquelle tous les cas d'un événement donné sont recueillis (que ces cas aient été vaccinés ou non). Dans ce cas oui, on s'attend à une répartition aléatoire de la survenue des cas par rapport aux dates de vaccination. Dans le cas de VAERS, ceci ne tient pas puisque la base rassemble des cas que des personnes (professionnels de santé ou non) ont signalés parce qu'ils leurs semblaient possible/probable que l'événement puisse être lié au vaccin, essentiellement parce que le délai

entre vaccination et événement leur semble évocateur. Les notifications sont en quelque sorte triées sur la base des délais d'apparition ce qui enlève tout sens à l'analyse que vous faites. Un peu comme si l'IFOP effectuait un sondage d'opinion dans la population d'un meeting de soutien à un candidat à la présidentielle. Il y a beaucoup de travaux là dessus, il faut les respecter. Bien cordialement

Comme on le voit, Bernard Bégaud n'hésite pas à qualifier de "non sens" une étude qui ne fait qu'appliquer des méthodes standards (comme nous le verrons dans la suite) de statistiques, mathématiques et probabilistes. Passant sur le ridicule de l'expression "Si et seulement si" (que Bernard Bégaud met lui-même en lettres capitales alors qu'il est toujours déconseillé d'utiliser un tel vocabulaire devant un mathématicien), on voit surtout que Bernard Bégaud est ***incapable*** de formaliser la notion de biais d'échantillon qu'il essaye pourtant de convoquer, sans qu'aucune formalisation ne soit invoquée, finissant par tomber dans une analogie dont on ne voit pas très bien le lien avec son propos. En réalité, il est toujours sans risque de prétendre qu'un échantillon est biaisé. En effet si l'on se résout à travailler sur des échantillons, c'est qu'on n'a pas accès à l'ensemble lui-même, mais seulement à une partie. Dès lors, les lois de probabilités sont effectivement connues sur l'échantillon, mais jamais sur l'ensemble lui-même, ce qui fait que l'on peut toujours estimer - puisqu'on ne le saura jamais - que l'échantillon est biaisé. Il s'agit d'une forme de doute sceptique consistant à estimer que si la vérité existe, on n'y aura jamais accès. Comme nous y reviendrons en conclusion, le doute sceptique n'est pas le doutes scientifique (cartésien) et s'y oppose même de manière frontale.

Pour la réponse à la critique de Bernard Bégaud, nous renvoyons le lecteur à l'annexe .1 qui détaille les failles méthodologiques aberrantes de l'auteur, s'agissant de comprendre et de formuler mathématiquement une notion aussi simple que la notion de biais d'échantillonnage.

1.3 La nécessaire intervention de nouveaux moyens

Comme nous l'avons rappelé, la méthode d'imputabilité de Bernard Bégaud - qui s'impose à l'ensemble des CRPV françaises - ne permet absolument pas de faire remonter de façon quantitative rigoureuse les signaux d'alerte liés aux vaccins.

La méthode Challenge-Déchallenge-Rechallenge reste une méthode quantitative, sans réelle justification mathématique ou statistique, qui ne relève que d'un parti pris relativement arbitraire. Surtout, elle échoue systématiquement sur la causalité des injections dans les effets indésirables.

Il est donc nécessaire, dans le cas d'une expérience telle que nous sommes en train de la vivre, de sortir de la méthode Bégaud et d'exploiter les informations quantitatives disponibles. En réalité, la réticence de la pharmacovigilance française usuelle quant à l'utilisation les bases de données quantitatives (VAERS, EUDRAVIGIOLANCE) relève sans doute plus de son incapacité structurelle à la programmation et à l'utilisation des outils informatiques qu'à une réelle objection méthodologique. L'hostilité qui est faite aux mathématiciens et aux informaticiens - pourtant spécialisés dans la description des méthodes mathématiques et numériques d'exploitation des données et dans leur programmation informatiques rigoureuses - relève ainsi beaucoup plus d'une aigreur liée à des personnes qui se voient contester dans leur domaine par des chercheurs qu'ils ne peuvent plus suivre. De fait, si l'on nous reproche d'exploiter les bases de données officielles de surveillance des effets indésirables, ce n'est pas parce que cela serait une ineptie scientifique, mais simplement parce que nous amenons des outils et des méthodes que la pharmacovigilance usuelle n'a aucune capacité de suivre actuellement. En outre, en tant que chercheurs spécialisés dans les outils mathématiques et informatiques, nous sommes capables d'aller chercher des signaux de pharmacovigilance

qui échappent complètement à l'école des CRPV. Ainsi, dans l'opposition qui nous est faite, c'est moins de science que de sociologie dont il est question : une vieille garde dépassée par les événements ne veut pas lâcher ses prérogatives sur un sujet qu'elle considère comme sa chasse gardée.

Un tel conservatisme possède cependant des conséquences dramatiques dans la surveillance nécessaire d'une expérience génique encore en phase d'AMM conditionnelle, pour laquelle les phases expérimentales ne sont pas terminées.

En réalité, la pharmacovigilance française, conformément à l'influence de Bernard Bégaud a verrouillé à triple tour l'ensemble de la pharmacovigilance s'agissant de sa capacité à faire remonter les signaux à une vaccination de masse.

1. ***D'abord elle a imposé (d'une manière ou d'une autre) une méthode d'imputabilité structurellement incompatible avec la causalité vaccinale;***
2. ***ensuite elle a systématiquement dénigré les bases de données liées à la remontée des effets indésirables. Comme l'écrivait [2] Il est cependant impossible de quantifier valablement un risque à partir des données recueillies.***
3. ***finalement, et contrairement à d'autres pays, elle refuse de donner l'accès aux chercheurs des données de pharmacovigilance passive.***

De fait nous allons étudier les signaux de pharmacovigilance de trois manière différentes, qui ont été totalement abandonnée - de façon inexpliquée - par la pharmacovigilance et qui disent cependant beaucoup sur les risques liés à la vaccination :

1. d'abord dans la suspicion évidente et inquiétante des effets dits facilitant dans la transmission du virus, qui surviennent après les injections. Ces conclusions ont été formulées bien avant que l'on s'accorde sur cette idée, par l'étude des courbes de mortalité en corrélation avec l'avancement vaccinal du début de l'année 2021. Il s'agit là typiquement d'un exemple de ce que la pharmacovigilance usuelle ne peut absolument pas obtenir par les méthodes avec lesquelles elle tente de confisquer le débat.
2. Ensuite une comparaison rigoureuse des risques entre les effets néfastes liés à la maladie COVID et les effets indésirables liés aux vaccins. Cette partie repose sur l'étude de la base européenne EUDRAVIGILANCE (en ce qui concerne les données de vaccinations), ainsi que de la base de données COVerAGE-DB https://osf.io/mpwjq/wiki/home/ qui recense pour le Covid19 les cas / décès par age, sexe et pays : https://osf.io/9dsfk.
3. Pour finir par l'étude comparative, dans le VAERS, entre les effets indésirables liés aux vaccins contre le Covid19 sur une année (2021 - début 2022) et l'ensemble des autres vaccins (une trentaine) sur les 30 dernières années. Comme on pourra le voir, cette étude sur le VAERS confirme à la fois les sur-risques de certaines pathologies (myocardites, thromboses etc...) liées aux injections ARNm. Surtout cette étude comparative montre que les vaccins contre le Covid impliquent des facteurs d'échelles très supérieurs à ceux des vaccins usuels. De ce point de vue, on ne donne plus une balance bénéfice-risque entre le vaccin et la maladie, mais une comparaison des vaccins entre eux. Ainsi, on permet que soit évaluée la qualité de la nouvelle technologie par rapport aux anciennes.

2 Mortalité COVID liée aux injections : un signal d'alerte sur les effets facilitant

L'ensemble de cette étude est disponible en annexe. Elle est également exécutable à l'adresse URL suivante :

https://tinyurl.com/5fn646jk

Les calculs et les documents résultants de ceux-là sont exécutables par n'importe quelle personne ayant un compte Google Drive qui soit en outre synchronisé avec sont ordinateur. Cette technologique scientifique collaborative permet l'échange sécurisé des programmes et des données entre chercheurs.

2.1 Les données

Deux pays particulièrement intéressants se sont offerts à l'étude analytique : l'Angleterre et le Portugal. Entre la fin de l'année 2020 et le printemps 2021, ces pays ont connu des courbes de mortalité COVID quasi identiques dans leur forme. Les données sont chargées à l'adresse suivante :

https://covid.ourworldindata.org/data/owid-covid-data.csv

Pour les pays on s'intéresse alors à l'incidence de mortalité Covid par million de personnes :

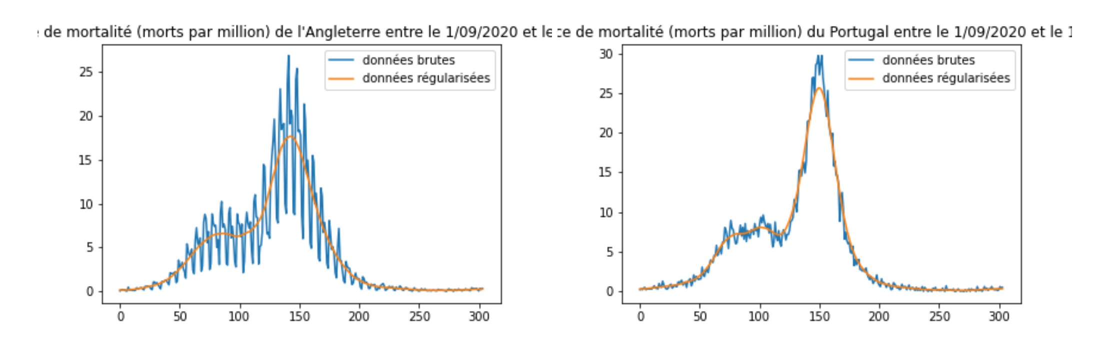


FIGURE 1 – mortalité en Angleterre et au Portugal

Après régularisation par convolution, on voit que les courbes - qui retranscrivent les incidences de mortalités entre le 1er septembre 2020 et le 1er Juillet 2021 - sont tout à fait similaires. Elles expriment 3 phases assez identiques :

1. Une montée automnale qui atteint un sommet et qui fait mine de redescendre
2. Un plateau qui s'étire sur plusieurs semaines
3. un point quasi anguleux qui annonce alors une montée violente et une courbe en cloche très caractéristique qui atteint un pic élevé avant de redescendre et se stabiliser sur une mortalité quai nulle

On peut superposer les graphes régularisés et voir que les similarités sont frappantes (voir figure 2) :

2.2 problématique et modélisation des données

A partir de leur "point anguleux" (augmentation brusque de l'incidence) les courbes de mortalités prennent une forme particulièrement régulière, typique de certaines courbes mathématiques. Il semble ainsi que le plateau que l'on observe soit le résultat d'un équilibre entre deux phénomènes :

1. une première épidémie (E) ayant atteint son pic et qui doit devenir décroissante (courbe en cloche) ;
2. un second phénomène (P) qui croit assez rapidement, empêchant la décroissance de la mortalité liée à (E), qui finit par atteindre un sommet avant de décroître.

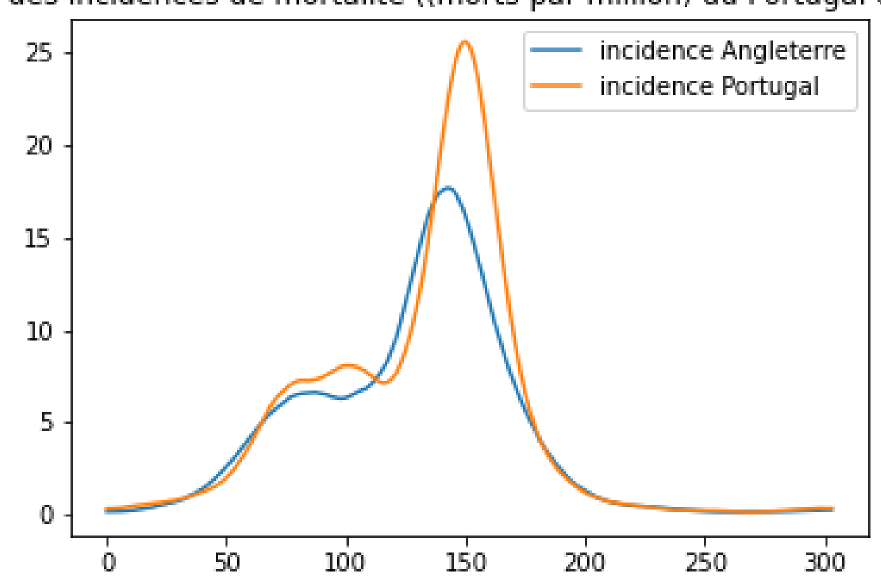


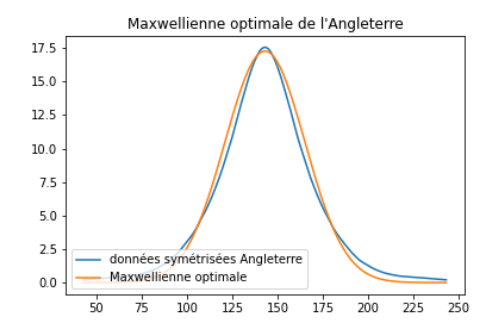
FIGURE 2 – comparaison des scénarios de mortalité

Il y a donc manifestement deux phénomènes superposés (E) et (P) pendant une certaine période, qui laisse ensuite le second phénomène (P) exister seul et s'éteindre. On part de l'hypothèse que le second phénomène (P) est seul observable à partir de son pic. Les questions son alors :

1. Peut-on modéliser la courbe à partir de son sommet jusqu'à son aplatissement final ?
2. si oui, peut-on remonter dans le passé à l'origine du phénomène (P) et que nous indique cette date ?

2.3 Modèle Maxwellien et estimation de la date d'origine

En exploitant la partie décroissante de la courbe à partir de son sommet, on est capable de montrer que l'incidence est modélisée par une courbe maxwellienne. Après avoir symétrisé la courbe par rapport à son sommet, on peut en effet retrouver simplement des paramètres maxwelliens qui permettent un fit remarquable :



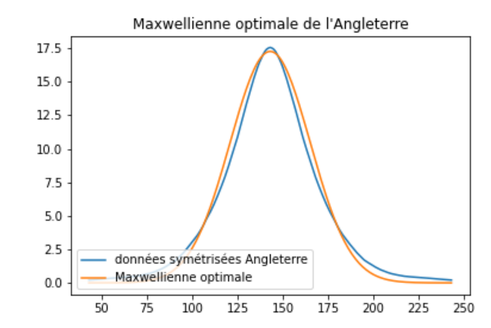
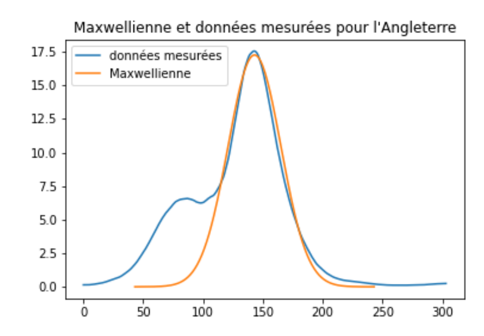


FIGURE 3 – Modèle Maxwellien

On est donc en capacité - sous l'hypothèse qu'un phénomène (P) s'est superposé à celui d'une épidémie décroissante - de modéliser le phénomène et donc de remonter à son origine temporelle.

Comme on a une courbe Maxwellienne, il est très simple de trouver l'origine : on peut la situer entre les bornes à 95% et 99% de l'intervalle de confiance. On arrive alors aux estimations suivantes :



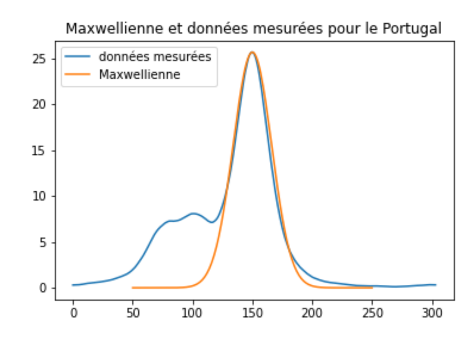


FIGURE 4 – Remontée dans le temps des Maxwelliennes

1. intervalle temporel correspondant à l'origine du phénomène maxwellien pour l'Angleterre : entre le 26 novembre 2020 et le 9 décembre 2020. Date de première vaccination : le 8 décembre 2020.
2. intervalle temporel correspondant à l'origine du phénomène maxwellien pour le Portugal : entre le 18 décembre 2020 et le 28 décembre 2020. Date de première vaccination : le 27 décembre 2020

Ainsi, d'un point de vue mathématique, il apparaît de façon très fiable qu'un phénomène de mortalité Maxwellien a manifestement pris son départ au moment où commençait la vaccination contre le COVID dans chacun des pays concernés. Comment expliquer un tel phénomène? La réponse probable est arrivée progressivement : il s'agissait sans doute d'un phénomène ADE (dits des anticorps facilitant). L'ADE est bien connue pour être un risque pour les infections à coronavirus, ainsi que pour la Dengue. L'ADE est une réponse immunitaire paradoxale qui rend une personne précédemment exposée à la maladie, ou à un vaccin la ciblant, plus – et non moins – vulnérable en cas d'infection ultérieure.

3 Pharmacovigilance européenne : construction d'une balance bénéfice-risque en fonction

3.1 Transparence et reproductibilité

L'ensemble du code informatique de traitement, d'analyse des données et calculs statistiques sont reproductibles et transparents. Ils peuvent être consultés à l'adresse suivante https://colab.research.google.com/drive/1-m-yTOSfSZcPONa5pmmkRYEM1OwT5iYf#scrollTo=lI8KYPfqV7nr. Les données ont été extraites le 4/04/2022.

Les données sont issues du site Europa mettant à disposition les cas recensés d'évènements indésirables graves et non graves dûs à au moins une injection du vaccin d'étude dans l'ensemble des pays européens : https://www.adrreports.eu/fr/search_subst.html#. Le détail du fonctionnement de la plateforme Eudravigilance peut être trouvé à l'adresse suivante : https://dumas.ccsd.cnrs.fr/dumas-01957354/document

3.2 Fiabilité des données

L'individuation des données est assurées par deux types de variables (i.e. ces variables prennent nécessairement des valeurs distinctes pour chaque individu)

— La première est la variable ICSR (individual Case Safety Report). Cette variable renvoie sur un lien URL dans lequel on peut retrouver les informations relatifs au remplissage de la pharmaco-vigilance.

— La seconde est la variable EU Local Number

Pour chaque personne, il est également possible de savoir si la remontée des EI a été réalisée par un personnel soignant ou un non soignant. La colonne "Primary Source Qualification" possède 2 valeurs différentes :

— "Healthcare Professionnal" : la déclaration des EI pour la personne concernée a été faite par un soignant;
— "Non Healthcare Professional" : cette déclaration a été faite par un non soignant.

Il est donc possible de voir en quelle proportion les remontées ont été réalisées.

3.3 Principe de comparaison pour l'élaboration d'une balance bénéfice-risque

Dans l'évaluation d'un vaccin qui doit protéger d'une maladie, il s'agit essentiellement de mettre en place la construction systématique d'une balance bénéfice-risque. Cette balance doit être construite de façon méthodique en fonction de chaque type de maladie et de chaque type de vaccin. Le principe général que nous posons ici est le suivant : vaccin ∼ maladie.

Cette transposition ainsi construite doit pouvoir être abordée de façon absolue et de façon relative :

1. De façon absolue : il s'agit de donner par exemple de définir des taux de mortalités et des taux de létalités pour les vaccins et pour les maladies, à la fois en population générale, par classe d'âge etc... Un des gros problèmes liés à cette comparaison est la suivante : on sait que les EI (Évènement Indésirable) ne font l'objet que d'une remontée incomplète. Ainsi, sur tous les indicateurs absolus liés à la pharmacovigilance, il faudra appliquer a priori un facteur multiplicatif représentant les effets de sous-déclaration. A l'inverse, on sait que les indicateurs absolus liés au Covid (en particulier les morts) peuvent être sur-déclarés (mort avec le Covid vs mort à cause du Covid). Il faudra donc toujours tenir compte de ces décalages dans les comparaisons.
2. De façon relative : il s'agit de la comparaison des catégories entre elles, autrement dit d'établir les facteurs de risque sur des critères donnés. Par exemple parmi les gens qui meurent (du vaccin ou de la maladie) on compare les facteurs de risque sur les classes d'âges ou sur les sexes. Pour ces comparaisons relatives et pour tenir compte du fait qu'il n'existe pas, pour le Covid, d'indicateur sur l'échelle de gravité.

Afin d'établir des indicateurs, nous estimons le sur-risque de présenter un effet grave du vaccin, d'hospitalisation et de décès en fonction du sexe et de l'âge en calculant des OR (Odd Ratio). Nous comparons ici le risque des hommes vis à vis des femmes et celles des différentes classes d'âge vis à vis de la population âgée entre 65 et 85 ans. Ce choix nous paraît cohérent au regard de la forte mortalité et surmortalité que le Covid19 a entraîné dans cette classe d'âge, l'une des plus touchée par cette maladie.

Nous calculons ensuite une balance bénéfice-risque en comparant les taux de létalité apparente du Covid19 (nombre de cas symptomatiques du Covid19 divisé par le nombre de décès du Covid19) à celui des vaccins ARNm Pfizer BioNTech et Moderna, le but ici étant de comparer la mortalité en cas de maladie du Covid19 à celle des vaccins en cas d'effets indésirables.

Nous présentons dans les sections qui suivent les résultats de ces calculs. Le lecteur pourra consulter l'annexe .2 afin d'obtenir les détails des calculs mathématiques.

	Total		Déclarations par personnel soignant	
	EI sérieux	EI non sérieux	EI sérieux	EI non sérieux
H	91104	142268	50651	39872
F	169410	400704	80736	99017
Sexe non spécifié	3441	5215	1153	2794
0-1mois	93	334	43	52
2mois-2ans	207	1154	56	590
3-11ans	1027	2564	583	781
12-17ans	11466	12378	6983	3282
18-64ans	165158	421499	73958	91462
65-85ans	70699	98492	39697	37605
≥ 86ans	15458	11766	11279	7911
Total	263955	548187	132540	141683

TABLE 1 – Nombre d'évènements indésirables sérieux et non sérieux en fonction du groupe d'âge et du sexe pour le vaccin Pfizer BioNTech

	Total		Déclarations par personnel soignant	
	EI sérieux	EI non sérieux	EI sérieux	EI non sérieux
H	38612	39599	26527	9111
F	61448	116645	34315	25195
Sexe non spécifié	1006	966	281	374
0-1mois	28	55	9	1
2mois-2ans	37	96	13	17
3-11ans	13	35	7	5
12-17ans	593	1127	436	518
18-64ans	72905	137689	39090	27853
65-85ans	23378	16908	17978	5451
≥ 86ans	4112	1300	3590	835
Total	101066	157210	61123	26480

TABLE 2 – Nombre d'évènements indésirables sérieux et non sérieux en fonction du groupe d'âge et du sexe pour le vaccin Moderna

3.4 Nombre d'évènements indésirables (EI) sérieux / non sérieux en fonction du sexe et du groupe d'âge

Les tables 1 et 2 montrent la répartition des évènements indésirables remontés dans la base de données européenne par sexe et classe d'âge.

— Sur l'ensemble des déclarations de cas graves (plus de 260000 cas) concernant le vaccin Pfizer, plus de 130000 ont été déclarées par un personnel soignant, soit environ 50% ;

— plus de 100000 déclarations de cas graves concernant le vaccin Moderna ont été rapportées dont plus de la moitié déclarée par un personnel soignant (60.4%).

3.5 Hospitalisations et décès par sexe et groupe d'âge

La table 3 montre la répartition des déclarations susceptibles d'avoir entraîné une hospitalisation ou un décès en fonction du sexe et du groupe d'âge selon le vaccin concerné (Pfizer BioNTech ou Moderna).

Nous pouvons constater qu'au prorata du nombre d'EI :

— les hommes sont plus touchés par des EI graves et les femmes par des EI non sérieux pour les deux vaccins ARNm ;

	Pfizer BioNTech		Moderna	
	Hospitalisation	Décès	Hospitalisations	Décès
H	28167	6034	15688	3339
F	35182	5513	16726	2301
Sexe non spécifié	535	115	221	48
0-1mois	29	14	13	3
2mois-2ans	35	4	9	2
3-11ans	238	9	3	1
12-17ans	3754	83	291	13
18-64ans	31795	1520	18642	1563
65-85ans	22036	5869	11297	2835
≥85ans	5997	4163	2380	1271
Total	63884	11662	32635	5688

TABLE 3 – Répartition des déclarations susceptibles d'avoir entraîné une hospitalisation ou un décès par sexe et groupe d'âge

	Pfizer BioNTech			Moderna		
	OR	lci	uci	OR	lci	uci
H vs F	1.51	1.49	1.52	1.85	1.81	1.88
0-1 mois	0.38	0.31	0.49	0.36	0.23	0.58
2 mois - 2 ans	0.25	0.21	0.29	0.27	0.19	0.40
3-11 ans	0.56	0.52	0.60	0.26	0.14	0.50
12-17 ans	1.29	1.25	1.33	0.38	0.34	0.42
18-64 ans	0.54	0.54	0.55	0.38	0.37	0.39
≥85 ans	1.83	1.78	1.87	2.28	2.14	2.44

TABLE 4 – Sur-risque de présenter un effet grave en cas d'effet indésirable par sexe et classe d'âge pour les vaccins ARNm Pfizer BioNTech et Moderna

— de la même façon, les adolescents âgés de 12 à 17 ans apparaissent plus touchés également par des EI grave, ainsi que les personnes âgées de plus 85 ans.

3.6 Pourcentage de risque de présenter un évènement indésirable grave en fonction du sexe et de la classe d'âge

La table 4 et la figure 5 présentent les OR (Odd Ratio) traduisant le sur-risque de sévérité des hommes comparé aux femmes et celui de chaque classe d'âge comparé au groupe d'âge des 65-85 ans. A partir de ces OR, il est possible de calculer la probabilité pour chaque population donnée d'être touché par un évènement indésirable grave en cas d'évènement indésirable. Nous pouvons remarquer que :

— les hommes présentent un sur-risque non négligeable de développer un effet indésirable grave par rapport aux femmes. Pour le vaccin Pfizer BioNTech, ils ont 40% de risque de développer un effet grave en cas d'effet indésirable contre 30% chez les femmes;

— les adolescents âgés de 12 à 17 ans présentent également un sur-risque de présenter un effet indésirable grave en cas d'effet indésirable comparé aux adultes de 65 à 85 ans, les plus enclins à développer des formes graves. En cas d'évènement indésirable, ils ont quasiment 1 risque sur 2 que cet effet soit grave. Cela est totalement contradictoire avec le fait que les formes graves du Covid19 sont rares dans cette classe d'âge avec des taux de létalité très faibles.

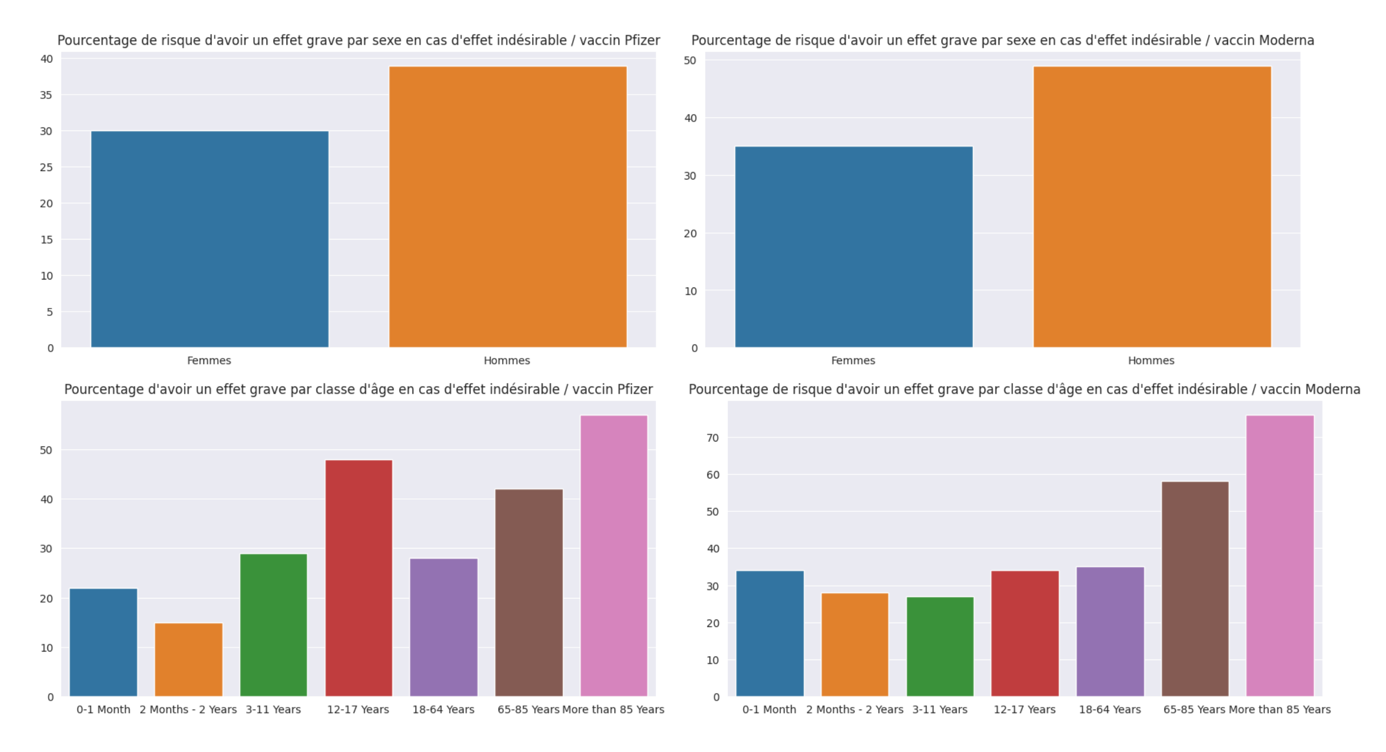


FIGURE 5 – Pourcentage de risque par sexe et classe d'âge de présenter un EI (Evènement indésirable) grave en cas d'EI pour les vaccins ARNm Pfizer BioNTech (à gauche) et Moderna (à droite)

	Pfizer BioNTech			Moderna		
	OR	lci	uci	OR	lci	uci
H vs F	2.08	2.05	2.12	2.42	2.36	2.47
0-1 mois	0.48	0.33	0.70	0.47	0.26	0.86
2 mois - 2 ans	0.17	0.12	0.24	0.18	0.09	0.36
3-11 ans	0.47	0.41	0.54	0.17	0.05	0.55
12-17 ans	1.24	1.20	1.29	0.52	0.45	0.59
18-64 ans	0.38	0.37	0.38	0.24	0.24	0.25
≥85 ans	1.88	1.82	1.94	2.01	1.90	2.13

TABLE 5 – Sur-risque d'hospitalisation par sexe et classe d'âge pour les vaccins ARNm Pfizer BioNTech et Moderna

3.7 Pourcentage de risque d'hospitalisation en fonction du sexe et de la classe d'âge

La table 3 indique le nombre d'hospitalisations et de décès par sexe et classe d'âge en fonction du vaccin. Nous pouvons remarquer que le nombre de déclarations d'hospitalisations chez les femmes est supérieur à celui des hommes. En revanche, ramené sur l'ensemble des déclarations dans chacune de ces populations, les taux d'hospitalisation sont plus forts chez les hommes que chez les femmes (plus de déclarations concernent les femmes) : respectivement 6.17% chez les femmes en cas d'EI contre 12.07% chez les hommes.
Comme nous pouvons le constater (voir table 5 et figure 6)

— qu'il existe un sur-risque d'hospitalisation chez les hommes comparé aux femmes en cas de survenue d'effet indésirable;
— de façon similaire à la sévérité, la population des 12-17 ans présente un sur-risque d'hospitalisation comparé à la population âgée de 65 à 85 ans; leur risque de présenter un EI grave étant de plus de 15% en cas d'EI.

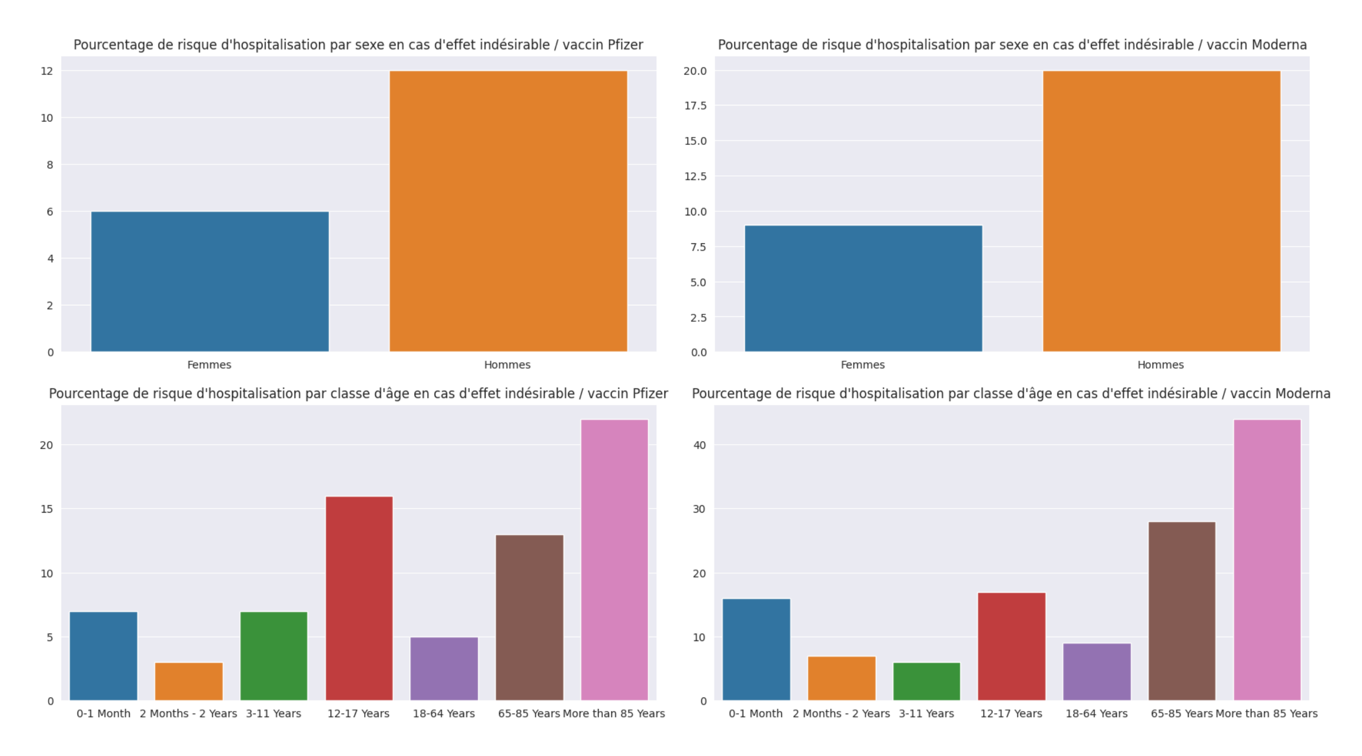


FIGURE 6 – Pourcentage de risque d'hospitalisation par sexe et classe d'âge pour les vaccins Pfizer BioNTech (à droite) et Moderna (à gauche)

	Pfizer			Moderna		
	OR	lci	uci	OR	lci	uci
H vs F	2.71	2.61	2.82	3.40	3.22	3.59
0-1 mois	0.94	0.55	1.60	0.49	0.15	1.56
2 mois - 2 ans	0.08	0.03	0.21	0.20	0.04	0.81
3-11 ans	0.06	0.03	0.13	0.28	0.03	2.03
12-17 ans	0.09	0.07	0.12	0.10	0.05	0.17
18-64 ans	0.07	0.06	0.07	0.09	0.09	0.10
≥85 ans	5.02	4.81	5.23	4.05	3.76	4.36

TABLE 6 – Sur-risque de décès par sexe et classe d'âge pour les vaccins ARNm Pfizer BioNTech et Moderna (la classe d'âge de référence étant celle des 65-85 ans)

3.8 Pourcentage de risque de décès en fonction du sexe et de la classe d'âge

Outre le fait que les hommes présentent un sur-risque d'hospitalisation pouvant être lié aux vaccins Covid19, ils présentent également un sur-risque de décès comparé aux femmes (voir table 6 et figure 7). Nous retrouvons également un très fort sur-risque de decès pouvant être lié aux vaccins ARNm Pfizer BioNTech et Moderna chez les personnes les plus âgées (≥85 ans) par rapport à celles âgées entre 65 et 85 ans et ce, de façon très significative. Il est à noter qu'il y a un probabilité non négligeable de décès en cas de survenue d'évènement indésirable chez les enfants et les adolescents pour le vaccin Moderna.

3.9 Comparaison avec le Covid19 : Calcul de la balance bénéfice-risque

Afin d'établir une balance bénéfice-risque concernant les vaccins Covid19, nous comparons les taux de létalité apparent du Covid19 à ceux des vaccins en cas d'effet indésirable sous la question : "Une personne a-t-elle plus de risque de mourir du Covid19 ou du vaccin si elle est victime d'effet indésirable ?".

Les données du Covid19 (nombre de cas et de décès par classe d'âge) sont issues de la base COVerAGE-DB(https://osf.io/mpwjq/wiki/home/, projet réunissant des centaines de chercheurs à travers l'Europe

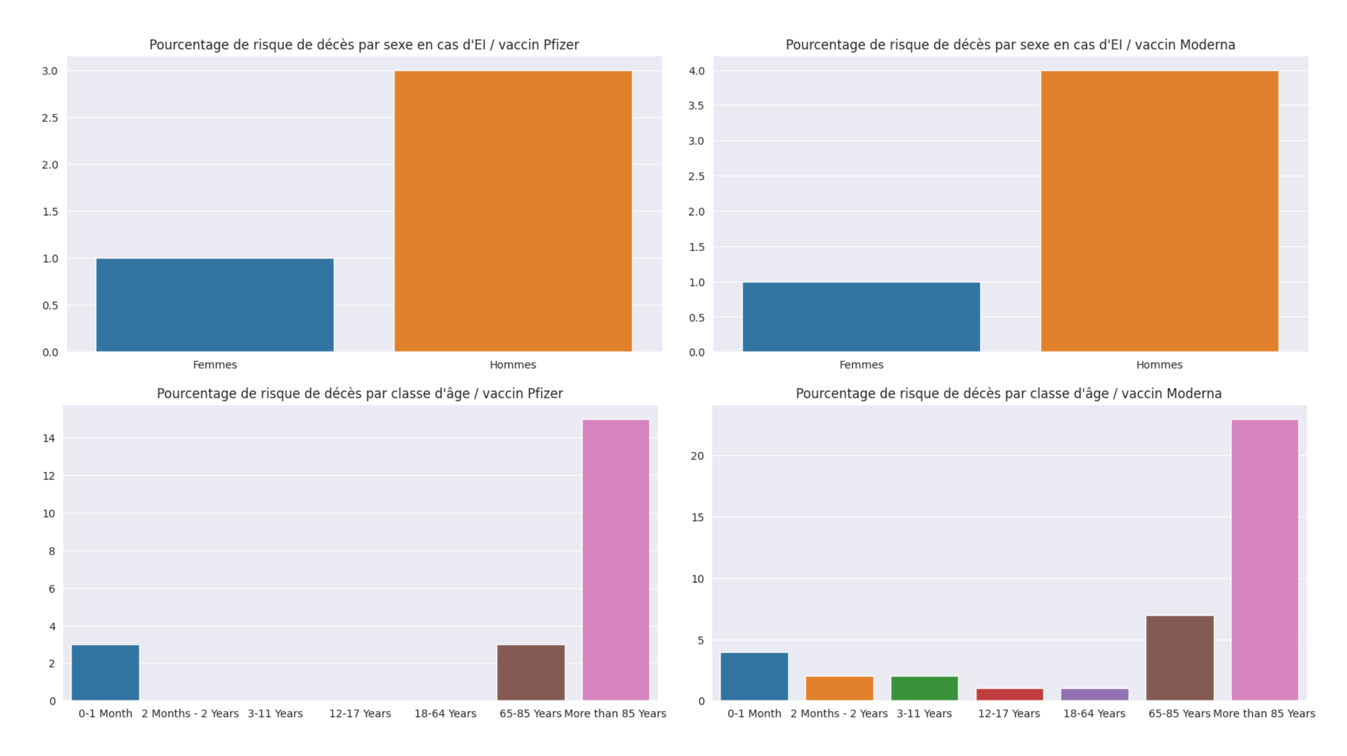


FIGURE 7 – Pourcentage de risque de décès par sexe et classe d'âge pour les vaccins ARNm Pfizer BioNTech (à gauche) et Moderna (à droite)

et dont les données quasi-quotidienne sont mise à jour depuis les données officielles des agences gouvernementales de chaque pays.

Les taux de létalité apparente (correspondant au nombre de cas connus / nombre de décès) pour la France depuis le 16.03.2020 jusqu'au 04.04.2022 sont visibles dans la table 7. Nous pouvons remarquer que les jeunes enfants et adolescents ont des taux de létalité du Covid19 extrêmement faibles (de l'ordre de 0.002%) alors que les personnes âgées ont malheureusement des taux de létalité beaucoup plus importants, notamment chez les plus de ≥90 ans.

Nous pouvons ainsi comparer pour chaque classe d'âge le taux de létalité apparent au regard de celui des vaccins ARNm en cas d'effet indésirable. La table 8 présentent ainsi les taux de létalité apparents (nombre de cas victimes d'évènement(s) indésirable(s) / nombre de décès) par sexe et classe d'âge.

Il apparaît clairement (voir figure 9 et 10) que la balance bénéfice-risque soit :

- en défaveur des injections ARNm chez les hommes;
- en défaveur pour l'ensemble des classes d'âges sauf pour celle des 65-85 ans.

Nous pouvons noter que les taux de létalité des injections ARNm sont très largement supérieurs à ceux de la maladie elle-même chez les jeunes, notamment avec un facteur très important pour les adolescents allant de 300 pour l'injection Pfizer BioNTech à plus de 500 pour l'injection Moderna(voir figure 11). La vaccination massive dans cette classe d'âge est donc de très loin plus néfaste que le Covid19 en l'absence de comorbidités ou de facteurs très aggravants.

	Cas	Décès	τ (en %)
F	5351199	40653	0.76
H	4689780	55367	1.18
0-9 ans	660126	13	0.002
10-19 ans	1427987	15	0.001
20-69 ans	7102691	16368	0.23
70-89 ans	779398	58645	7.52
≥ 90 ans	156813	21160	13.49

TABLE 7 – Taux de létalité apparent (noté τ) du Covid19 par sexe et par classe d'âge en France depuis le 16 mars 2020. Les données ont été extraites le 04/04/2022.

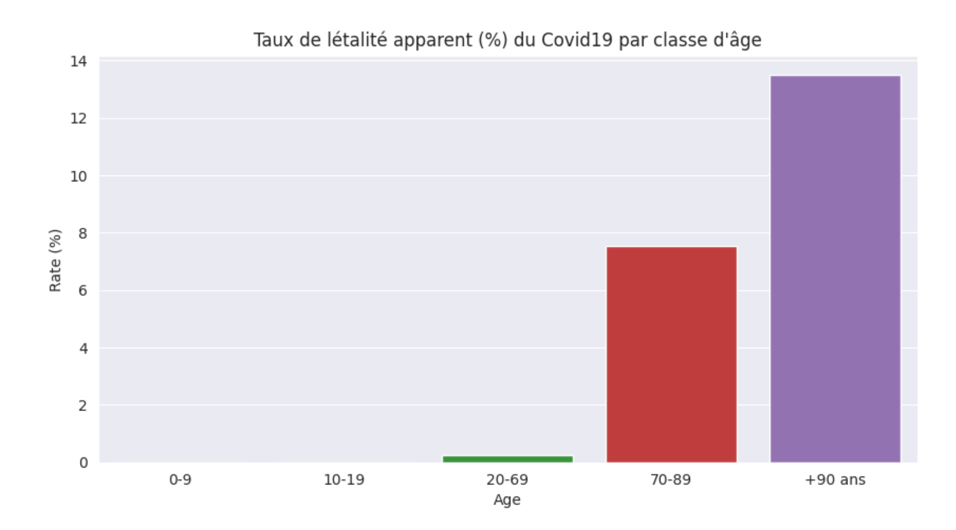


FIGURE 8 – Taux de létalité apparent du Covid19 par classe d'âge

4 Pharmacovigilance américaine : étude comparative de l'évolution temporelle des déclarations entre les période 1990-2020 et 2021-2022

4.1 Transparence et reproductibilité

L'ensemble du code informatique de traitement, d'analyse des données et calculs statistiques sont reproductibles et transparents. Ils peuvent être consultés à l'adresse suivante `https://colab.research.google.com/drive/18rCnV3DY19EtbFQvYbXorY-aPAcMAhyQ`.

L'ensemble des données brutes (toute année confondue) du système VAERS est téléchargeable sur ce

	τ_0 (%)	τ_1(%)	Moyenne
H	2.59	4.27	3.43
F	0.97	1.29	1.13
0-1 mois	3.28	3.61	6.89
2mois-2ans	0.29	1.5	0.89
3-11ans	0.25	2.08	1.16
12-17ans	0.35	0.76	0.55
18-64ans	0.26	0.74	0.5
65-85ans	3.47	7.04	5.25
≥85ans	15.29	23.48	19.35

TABLE 8 – Taux de létalité apparent des injections ARNm Pfizer BioNTech (noté τ_0) et Moderna (noté τ_1) par sexe et groupe d'âge

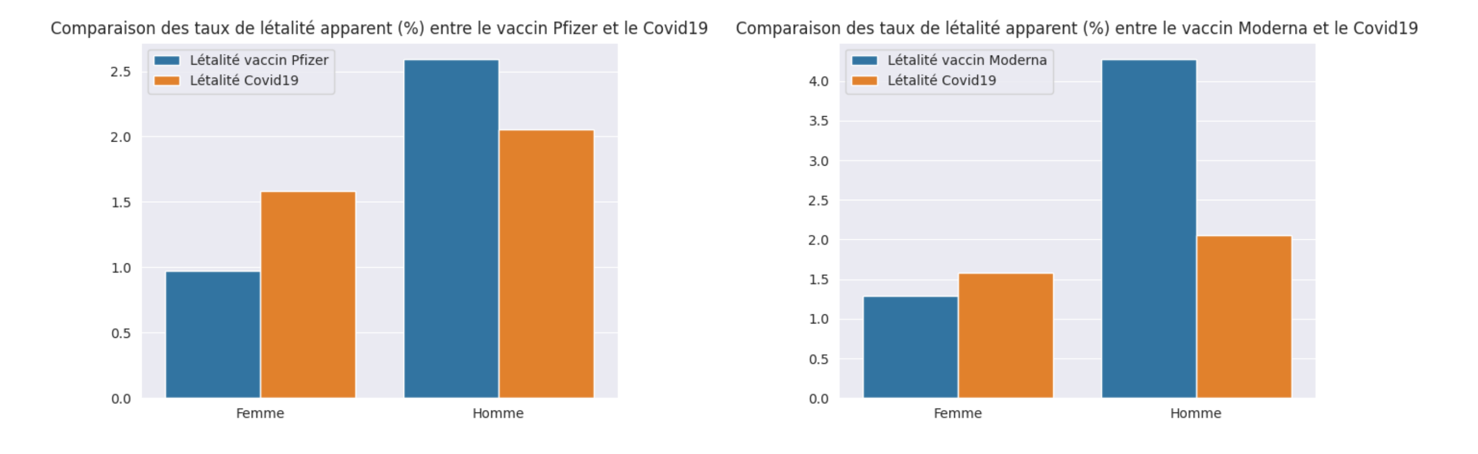


FIGURE 9 – Balance bénéfice-risque par sexe pour les injections ARNm Pfizer BioNTech (à gauche) et Moderna (à droite)

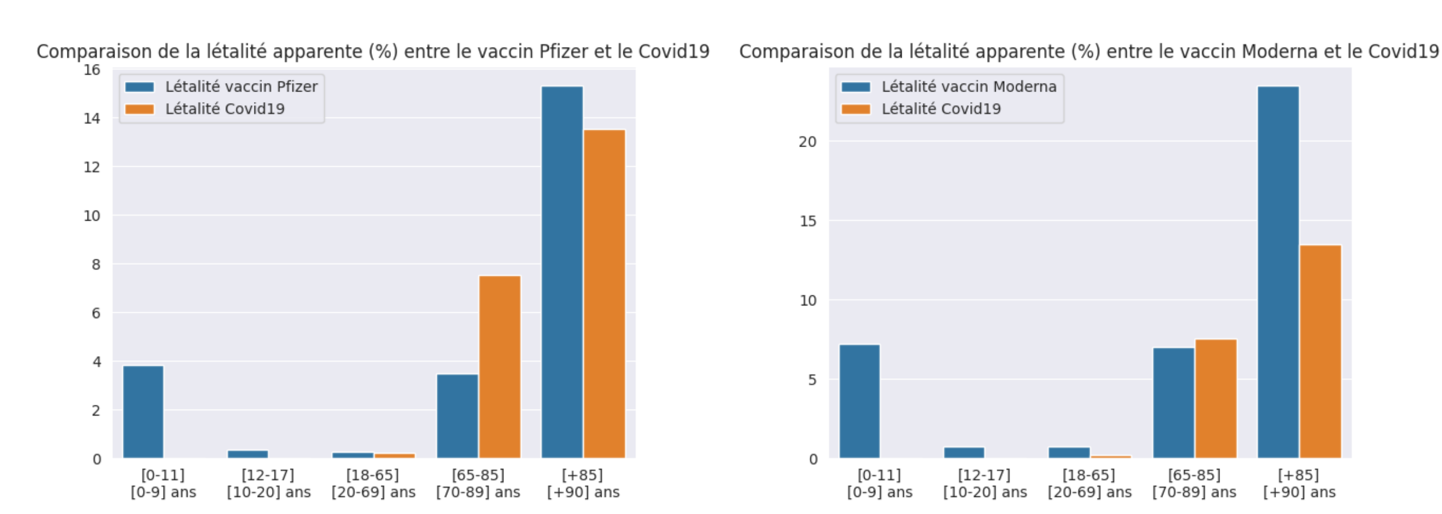


FIGURE 10 – Balance bénéfice-risque par classe d'âge pour les injections ARNm Pfizer BioNTech (à gauche) et Moderna (à droite)

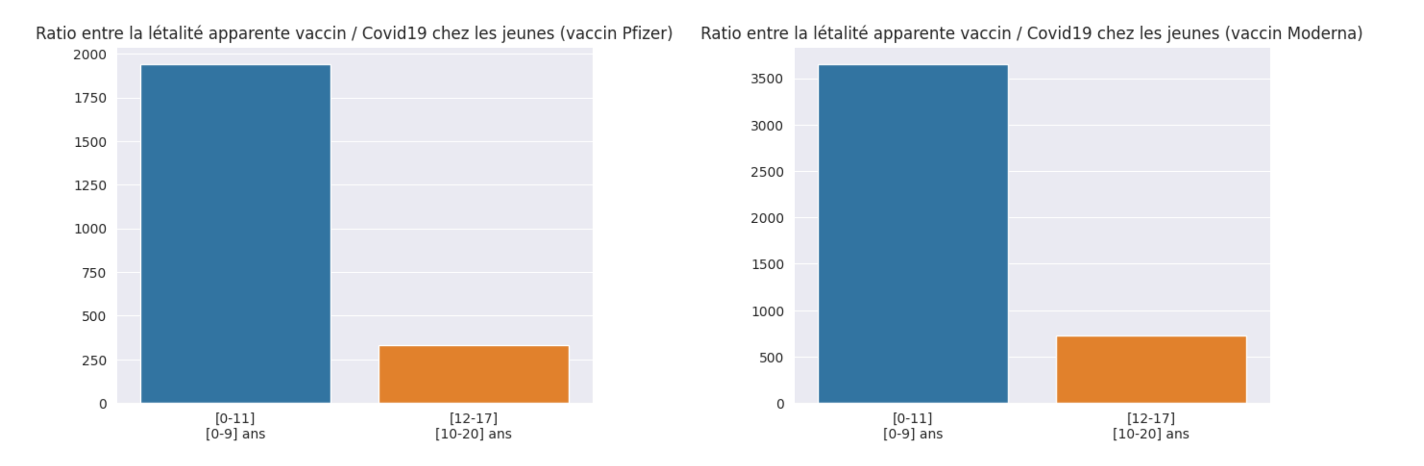


FIGURE 11 – Ratio des balances bénéfice-risque pour les classes d'âge jeunes concernant les injections Pfizer BioNTech (à gauche) et Moderna (à droite)

lien : `https://vaers.hhs.gov/eSubDownload/index.jsp?fn=AllVAERSDataCSVS.zip` Cette archive regroupe l'ensemble des déclarations d'évènements indésirables lié à l'administration d'une substance médicamenteuse depuis 1990 (un fichier de données par année).

Pour chaque année, nous avons ainsi 3 types de fichiers :

— un fichier décrivant les données personnelles pour chaque personne : identifiant de la personne, âge, sexe, explication textuelle du ou des symptômes ;
— un fichiers de déclarations de symptômes pour chaque personne : identifiant et symptômes (jusqu'à 5 symptômes)
— un fichier de déclaration mentionnant le vaccin susceptible d'être imputable aux effets indésirables décrits.

Ces fichiers peuvent tous être recoupés grâce à l'identifiant de la personne afin de connaître par exemple la temporalité de survenue de la pathologie déclarée vis-à-vis de sa date de vaccination en vue d'établir ou non une imputabilité post-vaccinale en se basant sur un critère chronologique par exemple.

4.2 Méthodologie

Notre méthodologie consiste à comparer la fréquence d'apparition d'une pathologie déclarée en 2021 et 2022 en la comparant avec les 30 dernières années cumulées tous vaccin confondu. Ceci permet d'établir un indicateur statistique pour la pathologie observée concernant les injections ARNm par rapport aux autres vaccins classiques, plus communément utilisés pendant ces 30 dernières années. Même si cette méthodologie s'appuie sur des valeurs absolues (non rapportées pour 100000 injections, à défaut de disposer des données précises concernant le nombre d'injections vaccinales par classe d'âge pour chaque année concernée), cette approche permet d'établir une comparaison indicative, permettant notamment de mettre en relief des signaux d'alerte.

Nous présentons dans ce qui suit le résultat de cette comparaison pour les hospitalisations, décès et quelques pathologies fréquentes rapportées dans le système de pharmacovigilance européen telles que les myocardites, péricardites, troubles menstruels, ainsi que des pathologies plus lourdes telles que les cécités, les accidents cardio-vasculaire (AVC) et les chocs anaphylactiques.

4.3 Comparaison des hospitalisations et décès par sexe et classe d'âge en 2021 et 2022 comparé à la période [1990-2020]

Durant la période 2021-2022, on décompte plus de 830000 déclarations dans la base de données du VAERS contre environ 700000 durant la période 1990-2020. La vaccination de masse aux États-Unis a bien entendu joué un rôle mais il est à noter que plus de 8% de ces déclarations concerne des décès ou des hospitalisations.

Nous pouvons remarquer, de la même façon depuis les données européennes, les déclarations de décès concernent légèrement plus les hommes que les femmes (cf table 9). Le nombre de déclarations d'hospitalisations et de décès est également plus important pour la période 2021-2022 que la période 1990-2020. On note ainsi une augmentation des hospitalisations déclarées dans le VAERS de plus de 78% en une seule année, ainsi qu'une augmentation d'environ 170% concernant les décès.

4.4 Myocardites et péricardites par sexe et classe d'âge en 2021-2022 comparé à la période 1990-2020

Le nombre de déclaration de myocardites a augmenté de façon très significative : 35 fois plus en 2021-2022 que durant la période 1990-2020 chez les 12-17 ans et 5 fois plus pour les 18-39 ans. Au niveau de l'ensemble des classe d'âge, on note ainsi une variation d'environ 80% en 2021-2022 en une seule année,

	Hospitalisations		Décès	
	1990-2020	2021-2022	1990-2020	2021-2022
H	16012	27956	2355	6238
F	17138	29665	1760	4702
0-4 ans	14014	190	2654	38
5-11 ans	1889	244	102	8
12-17 ans	2277	1678	123	52
18-39ans	5177	7755	199	398
40-65ans	5701	20029	373	2555
66-85ans	3977	23031	563	5720
≥ 85ans	528	5788	192	2574
Total	32963	58715	4206	11345

TABLE 9 – Nombre d'hospitalisations et de décès déclaré dans le VAERS sur les périodes 1990-2020 et 2021-2022

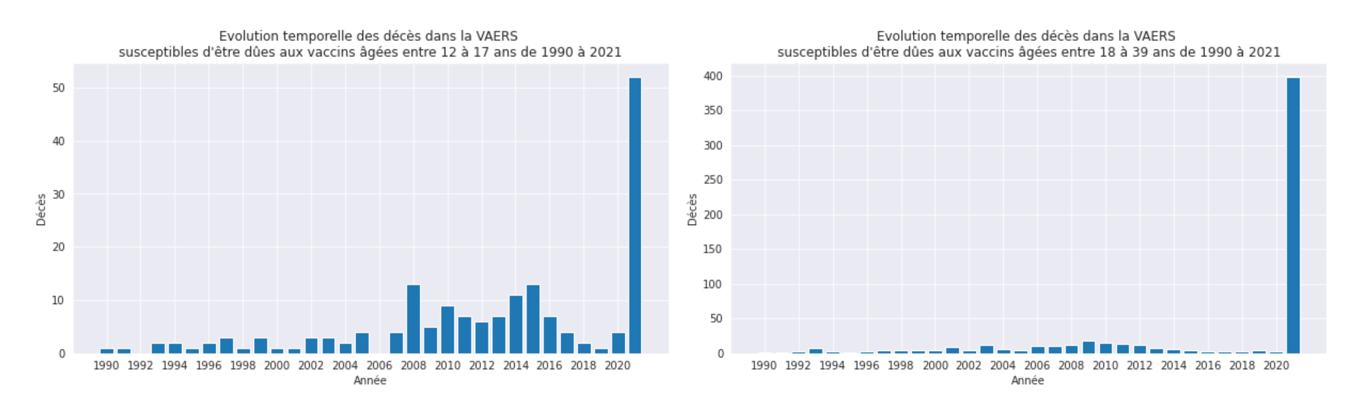


FIGURE 12 – Comparaison des déclarations de décès dans le VAERS entre les période par année. La dernière année correspond à 2021 et début 2022 cumulée.

phénomène visible dans toutes les classes d'âge.
Plus d'un tiers des déclarations de myocardites ont été réalisées par un personnel soignant (227 déclarations pour les 12-17 ans par exemple). Les 2/3 tiers de ces déclarations sont associées également à une hospitalisation. Elles concernent majoritairement les garçons (plus de 80%) quelle que soit la classe d'âge (497 chez les garçons contre 66 pour les filles chez les 12-17 ans par exemple). Ces déclarations sont remontées majoritairement entre 4 et 7 jours suivant la vaccination, ce qui laisse présumer une forte imputabilité liée à une injection ARNm (cf table 10). Ce résultat est en cohérence avec l'étude Epi-Phare [6] qui estime un sur-risque de myocardites chez les hommes de moins de 30 ans lié aux injections ARNm.

Nous pouvons remarquer le ratio entre le cumul des myocardites en 2021-2022 et celui sur la période 1990-2020 diminue avec l'âge avec un maximum pour la classe d'âge des 12-17 ans (cf Figure 14). Ceci laisse supposer que les jeunes sont plus enclins à être touchés par cette pathologie dans les 7 jours qui suivent la vaccination.

Le même phénomène est visible concernant les péricardites avec une augmentation de plus de 90% chez les 12-17 ans, environ 70% pour les 18-39 ans en 2021 et 2022 comparé à l'ensemble des années précédentes tous vaccin confondu.

Les injections ARNm sont très majoritairement liées à ces déclarations : plus de 99% de ces déclarations concernent les injections Pfizer BioNTech et Moderna (cf Figure 18).

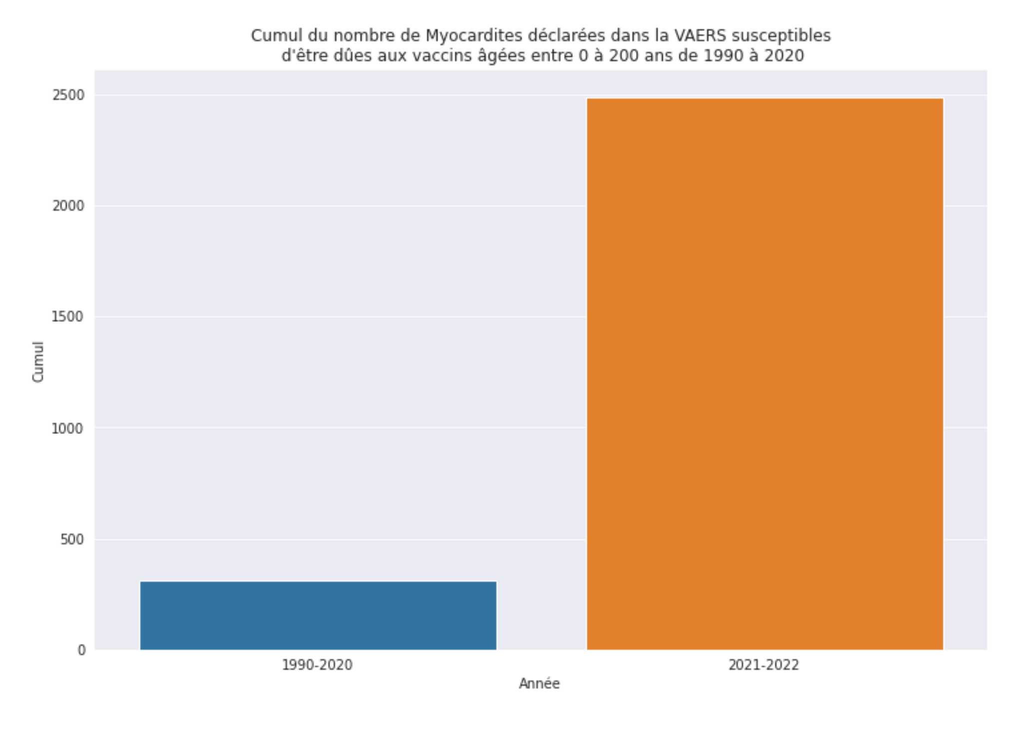


FIGURE 13 – Comparaison des déclarations de myocardites pour l'ensemble des classe d'âge entre les période 1990-2020 et 2021-2022

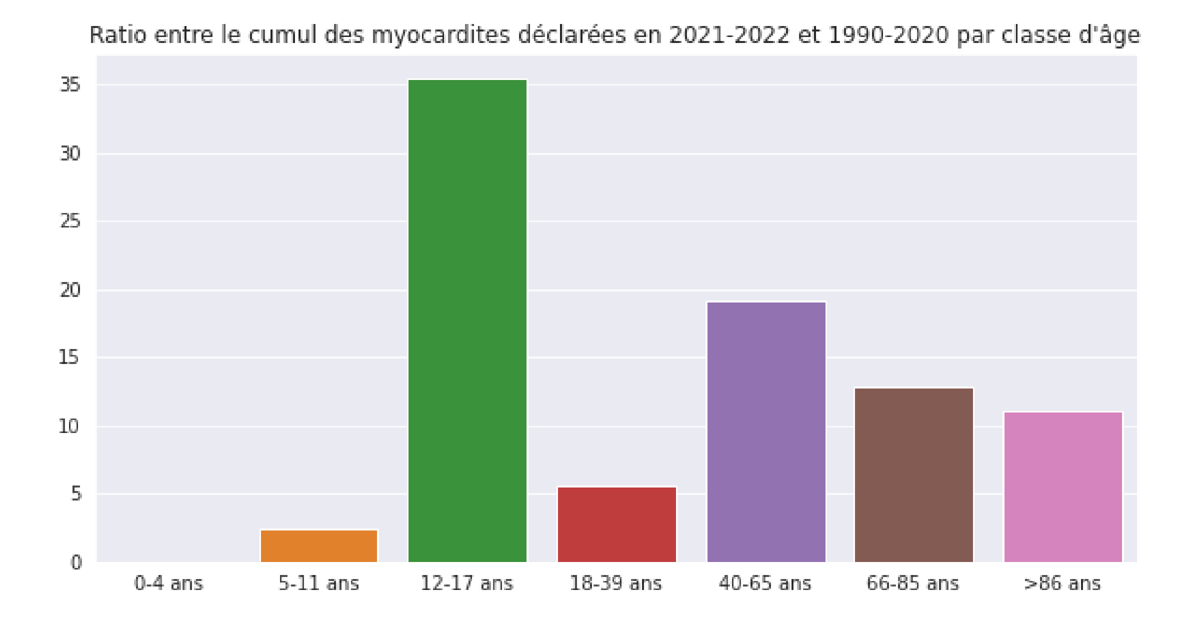


FIGURE 14 – Comparaison des ratios des déclarations de myocardites pour l'ensemble des classe d'âge entre les période 1990-2020 et 2021-2022

	Myocardites					Péricardites				
	J0	J1-J7	J8-J15	J15-J21	≥J21	J0	J1-J7	J8-J15	J15-J21	≥J21
0-4 ans	-	-	-	-	-	-	-	-	-	1
5-11 ans	-	5	2	2	3	-	3	-	-	5
12-17 ans	3	332	76	36	120	3	68	21	12	47
18-39ans	9	398	150	110	583	5	146	77	80	456
40-65ans	4	35	44	27	382	-	36	39	44	500
66-85ans	-	7	7	4	108	1	6	15	12	195
≥ 85ans	-	1	1	1	8	-	-	-	-	12
Total	16	778	280	180	1204	9	259	152	148	1216

TABLE 10 – Temporalité des déclarations en 2021 et 2022 des myocardites et péricardites dans le VAERS par sexe et classe d'âge

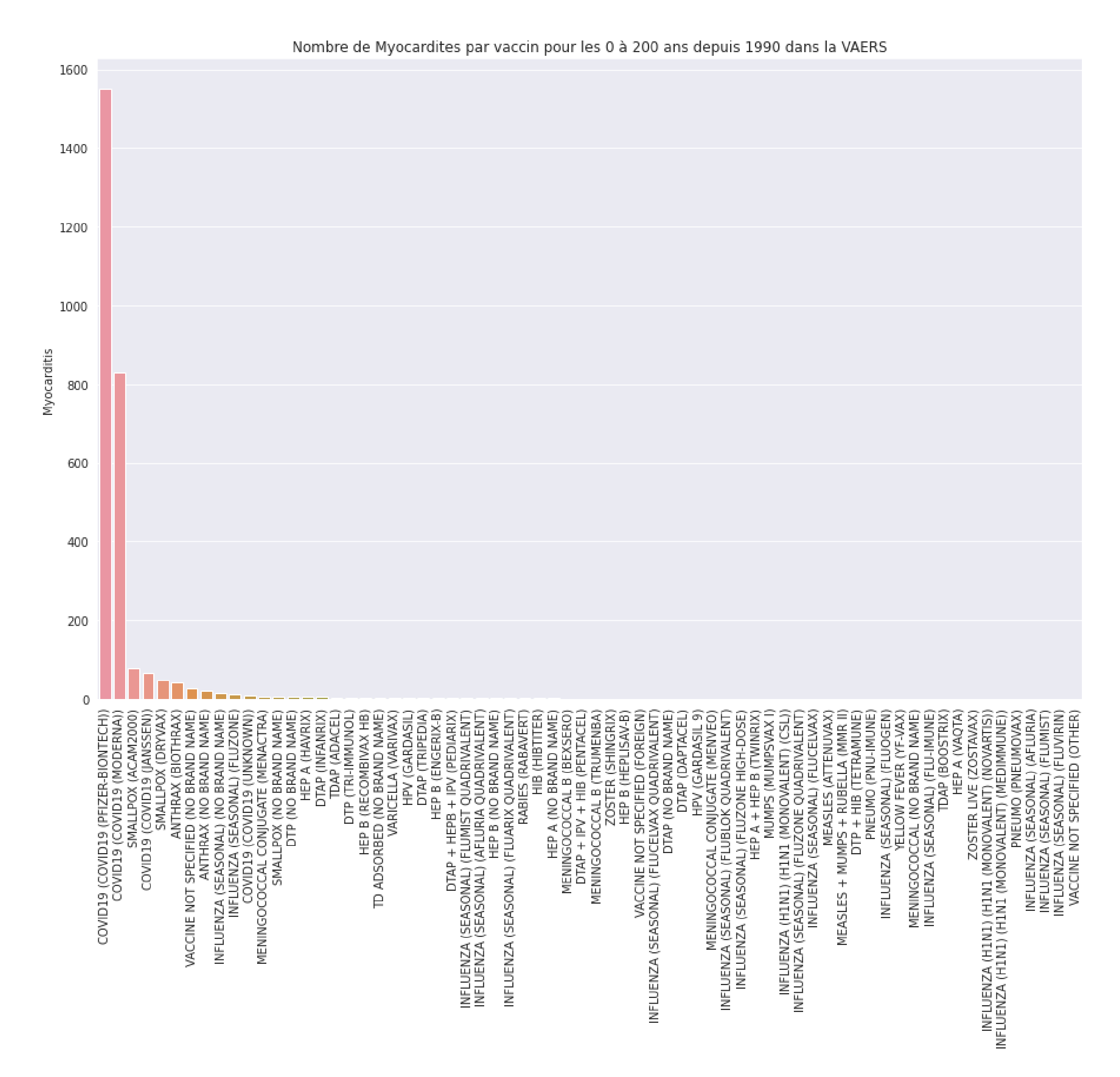


FIGURE 15 – Comparaison des déclarations de myocardites pour l'ensemble des classe d'âge par vaccin

	Myocardite		Péricardite		Trouble cycle menstruel		Choc anaphylac.		Cécité		AVC	
	1990-2020	2021-2022	1990-2020	2021-2022	1990-2020	2021-2022	1990-2020	2021-2022	1990-2020	2021-2022	1990-2020	2021-2022
H	267	1830	251	1082	-	-	449	340	190	405	166	1425
F	46	646	73	700	190	3588	698	1562	376	691	211	1714
0-4 ans	32	-	5	1	-	-	237	7	49	1	49	2
5-11 ans	5	12	2	8	-	-	166	7	51	12	11	1
12-17 ans	16	567	15	161	98	97	95	73	121	117	27	16
18-39ans	225	1272	224	762	78	2083	262	669	181	456	41	217
40-65ans	26	498	57	620	14	1408	309	963	124	325	101	1239
66-85ans	10	129	21	230	-	-	90	183	44	173	125	1350
≥ 85ans	1	11	2	12	-	-	2	6	3	27	23	314
Total	315	2489	326	1794	190	3588	1161	1908	573	1111	377	3139

TABLE 11 – Comparaison de l'évolution des déclarations des myocardites, péricardites, troubles du cycle menstruel, chocs anaphylactiques, cécités et AVC dans le VAERS sur les périodes 1990-2020 et 2021-2022

	Troubles menstruels				
	J0	J1-J7	J8-J15	J15-J21	≥J21
12-17 ans	1	8	2	4	77
18-39ans	45	239	149	185	1486
40-65ans	26	155	129	88	1003
Total	72	402	280	277	2566

TABLE 12 – Temporalité des déclarations des troubles menstruels en 2021 et 2022 dans le VAERS classe d'âge

4.5 Troubles du cycle menstruel en 2021-2022 comparé à la période 1990-2020

Les problèmes de troubles menstruels sont également très rapportés dans la base de données, avec une très forte augmentation, inédite dans le VAERS au regard des déclarations des années antérieures. En 2021-2022, on décompte ainsi 3588 déclarations contre 190 durant la période 1990-2020 et ainsi une variation de +1788% !

Ces troubles concernent les femmes jeunes mais également âgées de 40 à 65 ans. La très grande majorité de ces déclarations ont eu lieu après 21 jours post-vaccination, ce qui semble cohérent étant donné la nature de ce trouble.

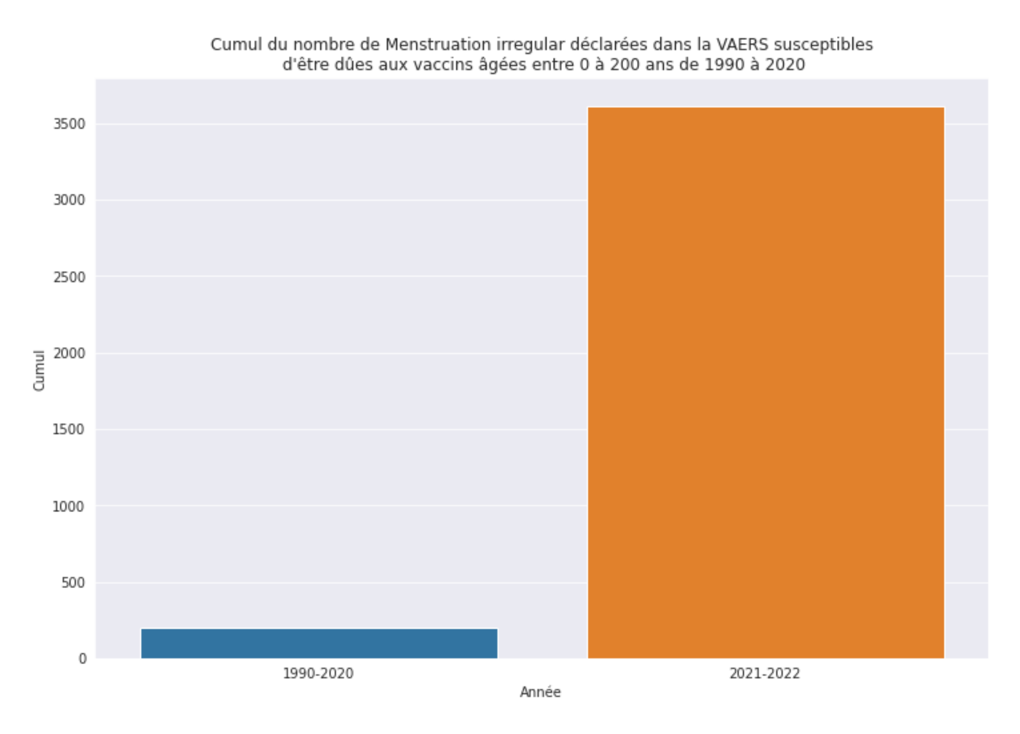


FIGURE 16 – Comparaison des déclarations des troubles du cycle menstruel entre les période 1990-2020 et 2021-2022 (toute classe d'âge confondue)

4.6 Chocs anaphylactiques par sexe et classe d'âge en 2021-2022 comparé à la période 1990-2020

Nous pouvons remarquer également une augmentation du nombre de déclarations des chocs anaphylactiques, notamment pour les personnes âgées entre 18 et 65 ans, dont environ 30% sont déclarées le même jour que l'injection et dont plus de la moitié des déclarations en 2021 et 2022 ont été remontées par un personnel soignant. Sur l'ensemble des classes d'âge, on note ainsi une variation de plus de 40% sur la

	Chocs anaphylactiques				
	J0	J1-J7	J8-J15	J15-J21	≥J21
0-4 ans	3	-	-	-	4
5-11 ans	1	3	2	-	1
12-17 ans	34	16	3	7	13
18-39ans	227	142	45	49	201
40-65ans	229	135	95	73	316
66-85ans	33	40	20	15	79
≥ 85ans	3	-	-	-	3
Total	530	336	165	144	617

TABLE 13 – Temporalité des déclarations des chocs anaphylactiques en 2021 et 2022 dans le VAERS classe d'âge

période 2021-2022, ce qui représente plus de 1900 cas contre 1161 entre 1990-2020 (cf Figure 17).

Ces chocs anaphylactiques concernent autant les hommes que les femmes : il n'y a pas de disparité au niveau du sexe dans les classes d'âge concernées. Par exemple, chez les 12-17 ans par exemple, environ la moité concerne les femmes et les hommes.

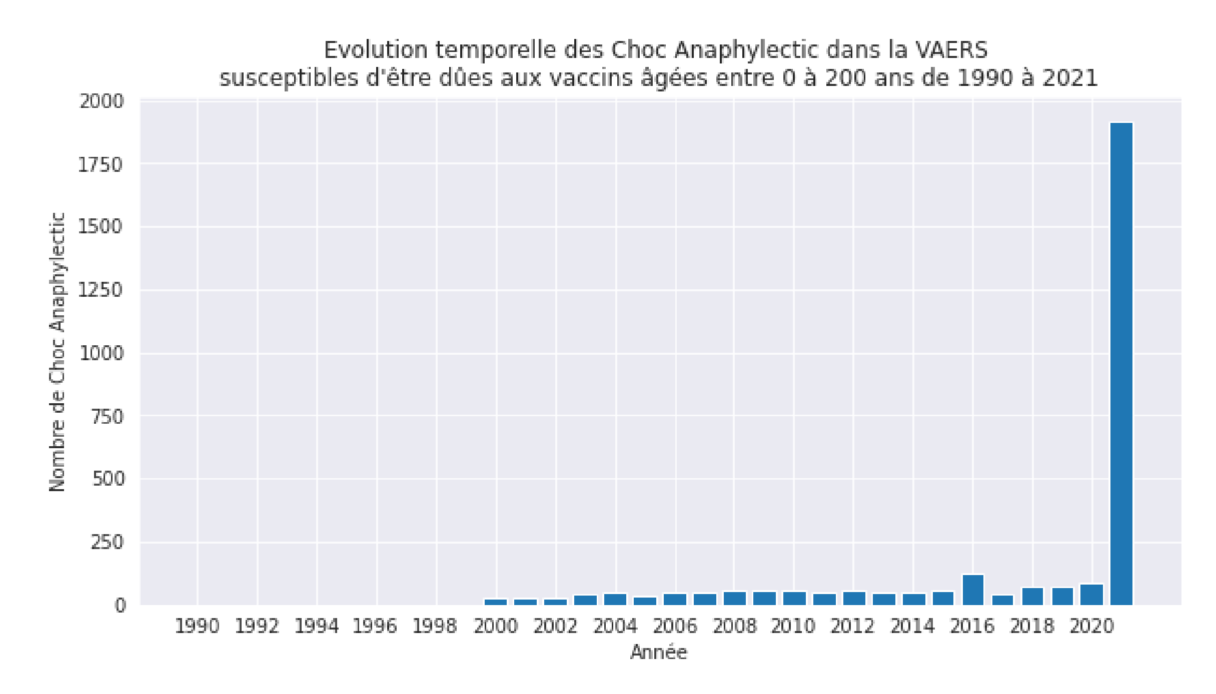


FIGURE 17 – Nombre de déclarations de chocs anaphylactiques par année toute classe d'âge confondue. La dernière année est le cumul des années 2021 et 2022.

4.7 Cécité par sexe et classe d'âge en 2021-2022 comparé à la période 1990-2020

Les déclarations de cécité ont également augmenté en 2021 et 2022 comparé aux années précédentes (voir table 14). Toutes les classes d'âge apparaissent touchées :

— chez les 12-17 ans, on remarque autant de déclarations en 2021 et 2022 que sur la période 1990-2020 (117 contre 121). La moitié des déclarations établies en 2021-2022 ont été faites par un personnel soignant (68 sur 117 au total). Dans cette classe d'âge, les déclarations ont été majoritairement remontées le jour même de l'injection (environ 62%) ;

— le nombre de cécité en 2021-2022 a dépassé celui sur la période 1990-2020 dans plusieurs classes d'âge, notamment celle des 18-39 ans et celle des 40-65 ans. Pour les 18-39 ans, on note une variation d'environ 60% dont plus d'un tiers des déclarations ont été remontées par un personnel soignant. Sur

	Cécité				
	J0	J1-J7	J8-J15	J15-J21	≥J21
0-4 ans	1	-	-	-	-
5-11 ans	5	4	2	1	-
12-17 ans	68	25	8	4	12
18-39ans	174	90	25	25	145
40-65ans	21	37	30	24	216
66-85ans	5	16	7	15	129
≥ 85ans	3	1	1	3	19
Total	277	173	73	72	512

TABLE 14 – Temporalité des déclarations des cécités en 2021 et 2022 dans le VAERS classe d'âge

l'ensemble des déclarations en 2021-2022, un tiers ont été faites le jour même de l'injection, ce qui laisse supposer une très forte imputabilité basée sur un critère chronologique à la vaccination ; les vaccins concernés étant les injections Pfizer BioNTech, Moderna mais également Janssen.

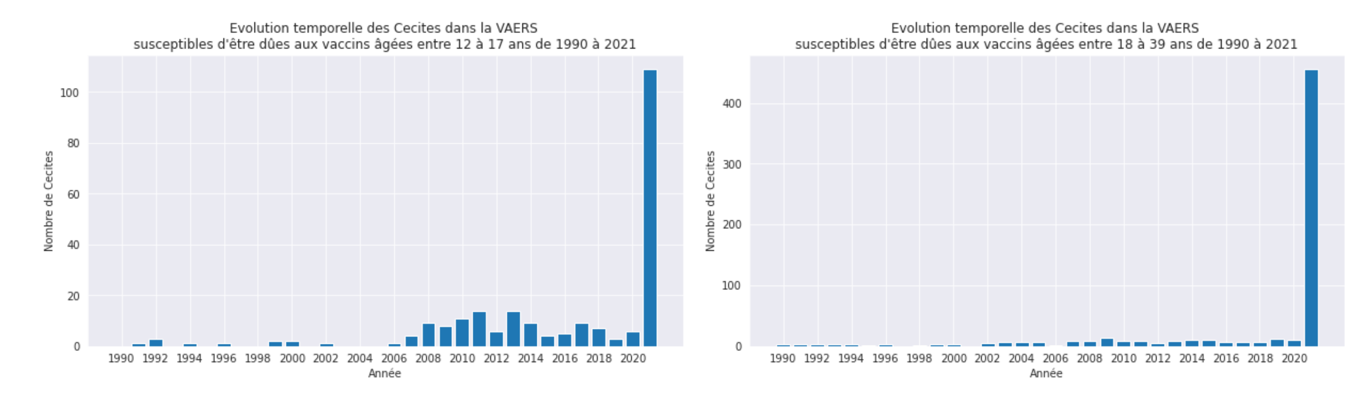


FIGURE 18 – Comparaison des déclarations de cécités entre les période 1990-2020 et 2021-2022. La dernière année représente le cumul des années 2021 et 2022.

4.8 Accident Cardio-Vasculaire par sexe et classe d'âge en 2021-2022 comparé à la période 1990-2020

Les AVC ont également fait l'objet d'une recrudescence de signalement dans le VAERS sur la période 2021-2022 : on note une variation de 732% comparé à la période 1990-2020 (voir Figure 19) ; les plus touchés étant les personnes âgées entre 66 et 85 ans. On remarque également que les adolescents sont touchés de façon inquiétante : en une seule année le nombre de signalement correspond à plus de la moitié cumulée sur la période 1990-2020.

En 2021, la large majorité de ces signalements ont été déclarés plus de 21 jours suivant la vaccination (voir table 15) mais il est à noter qu'une quantité non négligeable ont été déclarées entre 7 à 15 jours post-vaccination, ce qui laisse présumer une forte imputabilité des injections ARNm Covid19.

4.9 Discussion

Les statistiques détaillées dans les sections précédentes ne tiennent pas compte du phénomène de sous-déclaration. Très récemment, des travaux de recherche [5] estiment que le nombre de décès rapportés dans le VAERS est 20 fois en dessous de la réalité observée. Ce phénomène de sous-déclaration dans le système de pharmacovigilance "passif" est bien documenté dans la littérature scientifique. Nous renvoyons

	AVC				
	J0	J1-J7	J8-J15	J15-J21	≥J21
0-4 ans	-	-	-	-	2
5-11 ans	-	1	-	-	-
12-17 ans	2	2	1	-	9
18-39ans	7	14	16	16	160
40-65ans	12	84	94	107	924
66-85ans	13	89	98	93	1038
≥ 85ans	4	38	29	31	212
Total	8	24	61	62	764

TABLE 15 – Temporalité des déclarations d'AVC en 2021 et 2022 dans le VAERS classe d'âge

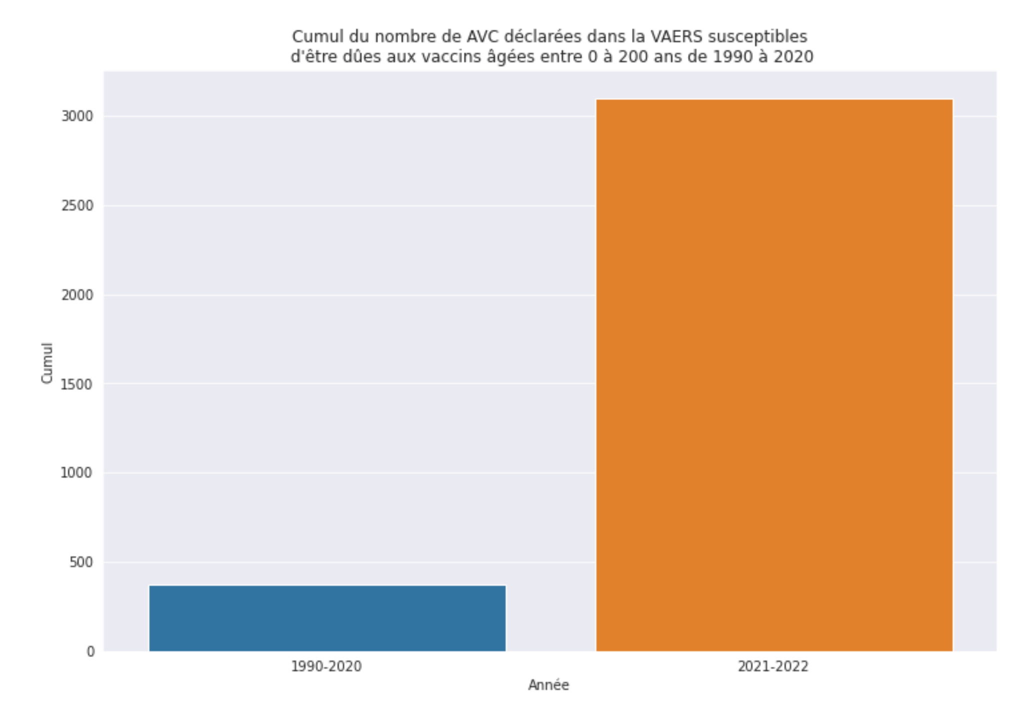


FIGURE 19 – Comparaison des déclarations d'AVC entre les période 1990-2020 et 2021-2022 (toute classe d'âge confondue)

le lecteur à consulter les références [1, 4, 3].

Les résultats présentés depuis le VAERS viennent également consolider ceux présentés depuis la base de données de pharmacovgilance européenne. En effet, en estimant la létalité pour chaque classe en considérant celle-ci comme étant le risque de décès en cas de symptôme du vaccin, il apparaît clairement que ces taux de létalité sont bien supérieurs à ceux du Covid19 pour les personnes âgées de moins de 65 ans.

Nous observons par ailleurs des évolutions de taux de létalité quasi-identiques dans les deux bases de données de pharmacovigilance (voir table 16 et 7) : les taux de létalité des injections ARNm augmentent significativement pour les personnes âgées de plus de 60 ans. Ce phénomène suggère que ces vaccins peuvent également nuire à des personnes avec un état de santé fragile (voir figure 20).

Les taux de létalité calculés depuis ces deux bases de données restent bien supérieurs à ceux du Covid19 pour les populations jeunes : il apparaît clairement que les injections ARNm sont plus létales que la maladie elle-même et qu'il n'y a pas de bénéfice à vacciner notamment les enfants, les adolescents ou les jeunes adultes, ne souffrant pas du Covid19 en l'absence de facteurs aggravants ou de comorbidités associées.

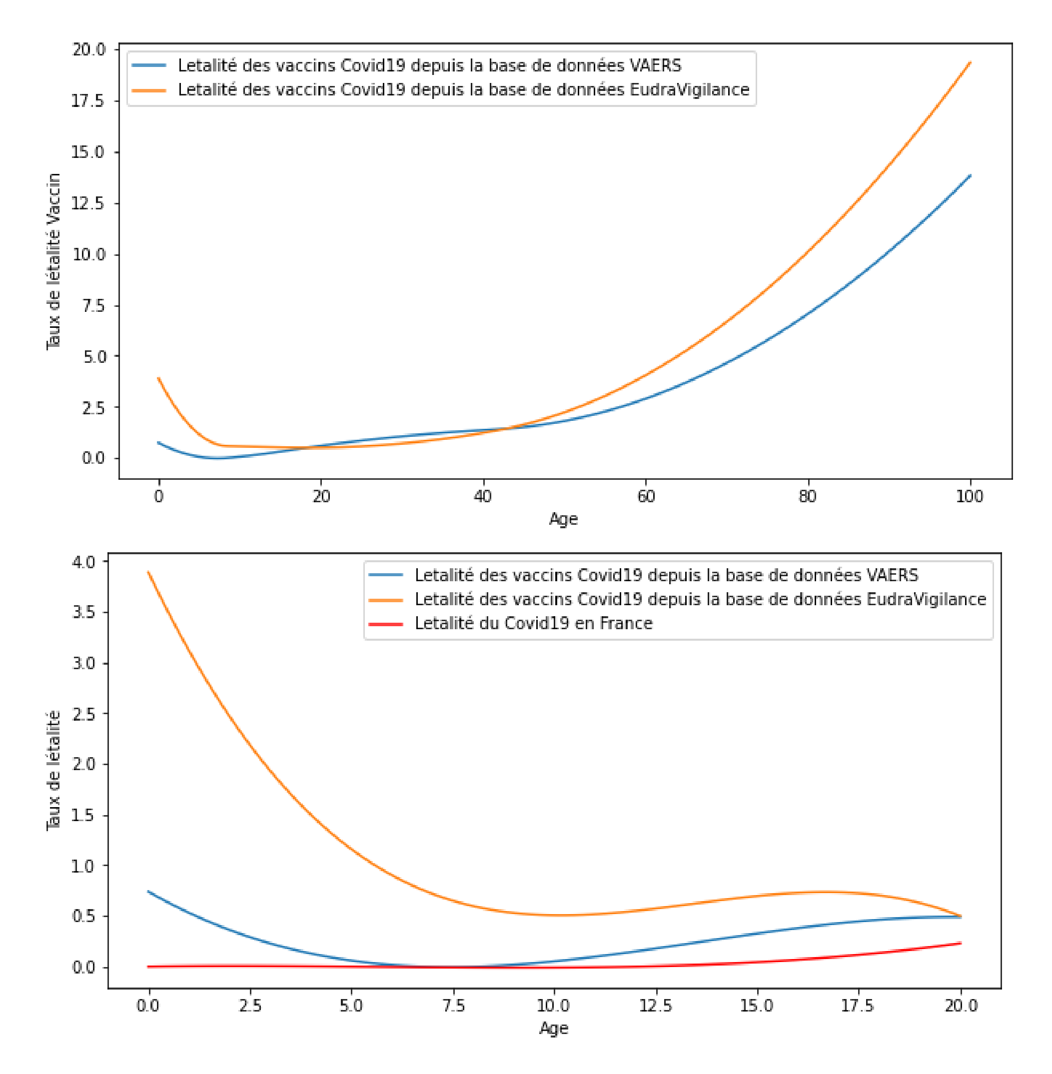


FIGURE 20 – Comparaison de l'évolution des taux de létalité en fonction de l'âge dans les bases de données de pharmacovigilance américaine (VAERS) et européenne (EudraVigilance).

De plus, la fréquence d'apparition de certaines pathologies comme les troubles du cycle menstruel, les cécités ou les AVC sont quasi-identiques dans les bases de données européennes et américaines (voir table 17). Cela témoigne que nous retrouvons la même proportion de signalements de ces pathologies parmi l'ensemble des signalements sur 2 continents différents et qui ont utilisé massivement les injections ARNm dans leur schémas vaccinaux.

5 Conclusion

5.1 Sur la déficience de la pharmacovigilance usuelle

Il apparaît de façon manifeste que la pharmacovigilance française n'a jamais été adaptée pour étudier sérieusement la question des effets indésirables concernant les vaccins :

1. Dans la méthode de Bégaud, qui domine dans les CRPV, les effets indésirables associés à un vaccin ont au mieux une imputabilité douteuse, dans la mesure où la chronologie déchallenge rechallenge ne peut avoir lieu
2. En outre, la méthode de Bégaud dédaigne le recours aux bases de données issues de la déclaration spontanée des effets indésirables

	Déclaration EI	Déclaration EI Décès	τ_3 (en %)
H	229311	6238	2.72
F	503597	4702	0.93
0-4 ans	5129	38	0.74
5-11 ans	11554	8	0.06
12-17 ans	33059	52	0.15
18-39ans	202029	398	0.19
40-65ans	319199	2555	0.80
66-85ans	154763	5720	3.69
≥ 85ans	18643	2574	13.81

TABLE 16 – Taux de létalité (noté τ_3) pouvant être imputables aux injections ARNm depuis le VAERS

	EUdraVigilance	VAERS
Myocardite	0.0085	0.003
Péricardite	0.0072	0.0021
Trouble du cycle menstruel	0.0045	0.0043
Choc anaphylactique	0.0071	0.0023
Cécité	0.0011	0.0013
AVC	0.0041	0.0039

TABLE 17 – Comparaison des fréquences d'apparition des myocardites, péricardites, troubles du cycle menstruel, choc anaphylactique, cécité et AVC parmi l'ensemble des signalements dans EUdraVigilance et le VAERS

Le débat est donc verrouillé à double tour : non seulement la méthode d'imputabilité ne permettra jamais de considérer des relations sérieuses de causalités, mais en outre on discrédite les seules sources de données qui pourraient honnêtement répondre aux questions de prudence.

Il est important de prendre en compte les effets catastrophiques du dogme de pharmacovigilance imposé depuis une petite cinquante d'années en France. Une fois que l'on a compris les traits essentiels de cette approche dominante, on peut également comprendre les réticences

5.2 Sur les évènements indésirables liés aux vaccins ARNm Covid19

Comme nous pouvons le constater, de nombreux indicateurs statistiques témoignent que les injections ARNm Pfizer BioNTech et Moderna peuvent être plus létales que les Covid19, notamment chez les jeunes où la balance bénéfice-risque est très clairement en défaveur du vaccin avec des facteurs de multiplication de létalité allant de 300 à plus de 500 par rapport à la létalité de la maladie elle-même.

La vaccination massive a entraîné également de nombreuses pathologies graves. Même si la causalité de la vaccination peut être discutée, les échelles de signalements en 2021-2022 par rapport à ceux de la période 1990-2020, notamment concernant les AVC et les troubles du cycle menstruel, sont véritablement inquiétantes et pourraient suggérer un véritable problème de santé publique sans précédent, bien plus important que la crise sanitaire que nous avons vécu durant ces deux dernières années.

Références

[1] R.A. ALATAWI Y.M.; Hansen. « Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event, Reporting System (VAERS) ». In : *Expert Opin. Drug Saf.* (2017), p. 761-767.

[2] Hugo ARLEGUI. « Les modèles quantitatifs pour l'évaluation du rapport bénéfice-risque des vaccins, état de l'art et axes d'amélioration ». thèse de k'Université de Bordeaux. 2019.

[3] S.A.W. HAZELL L.; Shakri. « Under-reporting of adverse drug reactions. A systematic review ». In : *Drug Saf.* 29 (2006), p. 385-396.

[4] C.L. MOORE T.J.; Bennett. « Underreporting of Hemorrhagic and Thrombotic Complications of Pharmaceuticals to the U.S. Food and Drug Administration: Empirical Findings for Warfarin, Clopidogrel, Ticlopidine, and Thalidomide from the Southern Network on Adverse Reactions (SONAR) ». In : *Thromb. Hemost* (2012), p. 905-907.

[5] H. PANTAZATOS P.; Seligmann. « COVID vaccination and age-stratified all-cause mortality risk ». In : (2021). URL : https://www.researchgate.net/publication/355581860_COVID_vaccination_and_age-stratified_all-cause_mortality_risk.

[6] Etude à partir des données du SYSTÈME NATIONAL DES DONNÉES DE SANTÉ. « Association entre les vaccins COVID-19 à ARN messager et la survenue de myocardite et péricardite chez les personnes de 12 à 50 ans en France ». In : (2021). URL : https://www.epi-phare.fr/rapports-detudes-et-publications/myocardite-pericardite-vaccination-covid19/.

[7] Marie THIMONNIER. « Olivier Véran peut-il affirmer qu'il n'y a aucun décès «imputable à la vaccination» contre le Covid-19 en France ? » In : *Libération* (2021).

6 Annexes

.1 Réflexions au sujet de la pharmacovigilance et la nécessité de la faire évoluer vers un travail partagé

Réflexions au sujet de la pharmacovigilance et la nécessité de la faire évoluer vers un travail partagé.

Vincent PAVAN, mathématicien, auteur, enseignant-chercheur, Aix-Marseille Université. Emmanuelle DARLES, informaticienne, enseignant-chercheur, Université de Poitiers.

La crise du COVID-19 constitue une période fascinante de notre époque qui donnera sans doute un travail titanesque aux futurs historiens devant se pencher sur la question. De façon évidente, le qualificatif qui décrit le mieux cette crise semble s'énoncer ainsi : total. Non seulement parce que cela touche l'ensemble de la planète, mais également parce que cela s'insinue dans tous les domaines d'activité, d'organisation, de pensée etc. des sociétés concernées. En première ligne et particulièrement exposé : le champ scientifique. Avec la pression invraisemblable mise sur les populations pour leur injecter de manière quasi forcée des substances prophylactiques encore en cours d'expérimentation, un sujet sensible – la pharmacovigilance – se mue carrément en enjeu démocratique capable à lui seul de faire tomber des gouvernements entiers. C'est dire si le droit de parole sur cet aspect devient lui-même un enjeu fondamental.

Comme tout domaine universitaire, celui de la pharmacovigilance possède sa propre structure sociale et ses propres portes parole autorisées. D'où cette question : qui peut légitimement prétendre parler au nom d'un savoir spécifique dans ce segment du réel ? La pharmacovigilance se définit selon les références officielles comme « la surveillance des médicaments et la prévention du risque d'effet indésirable résultant de leur utilisation, que ce risque soit potentiel ou avéré. ». D'une téléologie plutôt médicale, elle repose cependant sur l'utilisation certaine de la logique, des probabilités et des statistiques, employant le plus souvent des données en quantité considérables que l'on désigne aujourd'hui couramment sous le vocable de « Big Data ». Chasse gardée des quelques médecins ou pharmaciens devenus professeurs d'Université (eu égard à ses objectifs de santé publique), la discipline semble cependant ne pas résister – en ces temps de crises – à la venue de nouveaux entrants. Or, alors que cette supplémentation est appelée naturellement par les lacunes scientifiques criantes existant au sein de la pharmacovigilance usuelle, il apparaît – selon une règle sociologique classique – que cette mise à niveau est vécue comme une « intrusion » illégitime par les tenants de la parole autorisée. A ce titre, nous publions un mail qui nous fut adressé par un éminent spécialiste de pharmacovigilance, alors que nous avions rendue publique une étude sur les données de la VAERS qui avait été diffusée sur une page web, via la technologie « Google Colab » https://tinyurl.com/2p9fts38, et qui avait été présentée dans une vidéo du Conseil Scientifique Indépendant (CSI). Reprenons donc les propos de cette personne éminemment respectée, en l'occurrence Bernard Bégaud (Docteur en médecine, professeur à l'Université de Bordeaux).

« *Tout d'abord, la santé collective mérite un abord objectif reposant sur les faits et les données disponibles, sans orientation, tri ou interprétation a priori. Cela suppose aussi que l'on soit dans son domaine de compétence. La pharmacovigilance, la pharmaco-épidémiologie sont des spécialités dont la pratique demande formation et expérience, une longue expérience même si l'on ne veut pas tomber dans des pièges énormes. Je sais que cette crise nous a habitués à la négation des compétences et de la qualification scientifique (y compris chez les défenseurs des vaccins du reste) mais quand même ; restons humbles et rigoureux. Personnellement si je connais très bien, pour être à l'origine de quelques-unes, les méthodes de la pharmacovigilance et de la pharmaco-épidémiologie, je ne connais rien ou peu à l'immunologie ou à la génétique et je ne me permettrais pas de faire une démonstration ou un commentaire sur le vecteur viral utilisé dans les vaccins anti-COVID ou sur le caractère « naturel » ou non de telle séquence du SARS-VOV2.* »

Ainsi, comme on peut le constater sobrement, Il s'agissait bien sûr de nous dénier (poliment mais fermement) le droit – par des effets de pure autorité – de dire quoi que ce soit ayant pour thème la pharmacovigilance sans y avoir été invité par quelques personnes éminentes du champ considéré. Attitude très peu scientifique cependant puisque ce qui caractérise l'acceptabilité du discours en science ce n'est jamais la personne elle-même mais le contenu de son propos, et que les règles de la méthode ne sont jamais figées mais évoluent précisément sous la forme de « révolution scientifique » (Thomas Kuhn) lorsque les paradigmes du champ se trouvent brutalement transformés par l'apparition d'idées et de méthodes nouvelles. De plus la tendance à vouloir investir le champ de la pharmacovigilance par les tenants de la physique, des mathématiques, de l'informatique et de la logique est d'autant plus désirable que la notion de causalité à la base des considérations de cette discipline relève d'un discours général de l'épistémologie (laquelle cherche depuis toujours à formaliser ce concept de façon précise) qui ne saurait donc être le domaine privilégié de quelques personnes enfermées dans une tour d'ivoire universitaire. Ainsi, pour les tenants du formalisme en science, il apparaît de façon stupéfiante que la pharmacovigilance actuelle se situe en dehors des discours formels. Autrement dit, il s'agit manifestement bien plus d'une discussion qui manipule les « prénotions » (au sens de Durkheim) qu'une discipline ayant établi des « notions » d'ordre scientifique. Avec des conséquences inquiétantes dans pratique de la santé publique. C'est donc bien un manque de rigueur formelle sur les notions de pharmacovigilance que nous allons évoquer.

Ainsi commençons par étudier la légitimité dont se réclame l'éminent professeur en termes de pharmacovigilance lorsqu'il affirme que « Personnellement [si] je connais très bien, pour être à l'origine de quelques-unes, les méthodes de la pharmacovigilance et de la pharmaco-épidémiologie ». Après renseignement, on apprendra que l'auteur se réfère à l'une des méthodes qui constitue en France la référence d'imputabilité et dont il est l'un des instigateurs : le protocole dit « challenge-déchallenge-rechallenge ». Nous allons mener sur cette approche une étude en logique formelle qui va montrer qu'en réalité elle n'a pas la rigueur qu'elle prétend avoir et qu'il est très difficile de la considérer en tant que méthode scientifique. En pratique cette méthode (que nous abrégerons par CDR dans la suite) doit permettre de répondre à la question de savoir jusqu'où et comment la prise d'un médicament peut ou non devenir la cause d'un effet indésirable. Cette méthode fonctionne essentiellement en trois étapes :

1. On administre le médicament et l'effet indésirable survient;
2. On cesse l'administration du médicament et l'effet indésirable disparaît;
3. On administre à nouveau le médicament et l'effet indésirable réapparaît.

Dans quelle mesure cet enchaînement permet-il d'affirmer – ou pas - que le médicament est bien la cause de l'effet indésirable? Une manière de proposer une réponse à cette question consiste à formaliser – dans le langage de la logique des propositions - le déroulement du protocole et faire des calculs (logiques) à partir de cette formalisation. Cette petite réflexion va nous permettre de voir combien les choses les plus anodines en apparence sont cependant parfois les moins claires en théorie. Avant de poursuivre, avertissons notre lecteur : ce que nous allons écrire dans cette formalisation logique du CDR lui sera d'abord totalement incompréhensible avant que la conclusion finale que nous donnerons ne lui rende les choses totalement évidentes. En avant donc pour la formalisation du raisonnement sous-tendant le CDR.

Pour cela on note par A l'évènement associé à la prise du médicament et par B l'évènement associé à la survenue de l'effet indésirable. Par non(A) et non(B) on entend la négation des évènements précédents, c'est-à-dire l'absence de prise de médicament et l'absence de survenue de l'effet indésirable. L'imputabilité de l'effet indésirable par le vaccin se traduit – dans le langage de la logique formelle – par l'implication

A=>B : autrement dit, si on prend le médicament, alors l'effet indésirable apparaît. En quoi le protocole CDR constitue-t-il une preuve de l'imputabilité ? Autrement dit en quoi le fait d'associer entre elles certaines propositions liées aux événements A, B, non(A), on(B) peut-elle nous permettre d'affirmer une relation de cause à effet entre A et B ? Pour cela il faut immédiatement se poser la question de la notion de preuve en logique.

Définition 1 *Définition (méthode de preuve) : Dans le langage de propositions, une formule* $F(A,B)$ *construite sur les évènements A et B est une preuve de l'imputabilité* $[A \Rightarrow B]$ *lorsque la formule* $[F(A,B) \Rightarrow [A \Rightarrow B]]$ *est une tautologie.*

Bien sûr, tout cela constitue – au premier abord – un ésotérisme total pour qui n'a jamais pratiqué le calcul en logique de propositions. Mais s'agissant de justifier rigoureusement qu'une série d'observation permet effectivement de déduire une imputabilité, le minimum que l'on est en droit d'exiger c'est que les règles de déduction soient correctement respectées. Sans entrer dans les détails du calcul logique (ni dans la définition des objets manipulés et des concepts employés tel celui de tautologie) il apparaît la chose suivante : la formule $F(A,B) = [A\ et\ B]$ (c'est-à-dire j'ai pris le médicament et l'effet secondaire est apparu, autrement dit le challenge) ne constitue pas – à elle seule – une preuve de l'imputabilité. Autrement dit de façon strictement logique :

La formule $[[A\ et\ B] \Rightarrow [A \Rightarrow B]]$ n'est pas une tautologie.

De ce point de vue, on comprend aisément qu'il y ait besoin de compléter les informations liant les propositions A et B (et donc besoin « d'expérimenter » une nouvelle situation) si l'on veut « être sûr » de montrer la causalité entre A et B. C'est ce qui préside, manifestement, à l'établissement de l'étape de déchallenge dans le protocole CDR, c'est-à-dire à l'observation simultanée de non(A) (je ne prends pas de médicament) et de non(B) (je ne souffre pas d'effet indésirable). Du point de vue du calcul formel il apparaît alors effectivement le résultat suivant :

la formule $[[A\ et\ B]\ \ et\ \ [non(A)\ et\ non(B)] \Rightarrow (A \Rightarrow B)]$ est une tautologie

Ainsi, dans un premier temps, il semble que la séquence : j'ai pris le médicament, l'effet indésirable est apparu, puis j'ai arrêté le médicament et l'effet indésirable a disparu, soit une preuve de l'imputabilité. De cette façon, on pourrait être remarquablement satisfait de voir que le protocole répond de manière assez stricte à la notion de preuve en logique formelle. Et que l'on aurait assurément la place pour se féliciter de la rigueur de la méthode. Nous allons voir hélas qu'il n'en est rien et que la séquence CDR formalisée dans le cadre de la logique des propositions permet de tout affirmer (c'est-à-dire tout et son contraire). Remarquons au passage une première chose : dans la preuve CDR, l'étape de rechallenge est totalement inutile puisqu'elle constitue simplement une redondance de l'étape de challenge. Sa justification montre que l'on se situe déjà en dehors de l'aspect purement formel des choses. Mais il y a bien plus grave cependant du point de vue logique : c'est le fait que l'étape challenge-déchallenge permet aussi de montrer que la prise du médicament n'est pas la cause de l'effet indésirable. Du point de vue formel, on a ainsi :

la formule $[[A\ et\ B]\ \ et\ \ [non(A)\ et\ non(B)] \Rightarrow non(A \Rightarrow B)]$ est une tautologie

Pourquoi un tel résultat ? et bien simplement parce que dans la logique des propositions, la formule $[[A\ et\ B]\ \ et\ \ [non(A)\ \ et\ non(B)]]$ constitue ce qu'on appelle une absurdité et qu'elle donc peut servir à déduire n'importe quelle autre proposition du langage. A quelle réalité nous renvoie cette observation logique ? A la chose suivante : si vous avez pris un médicament (proposition A) alors vous ne pouvez plus affirmer ensuite que vous ne l'avez pas pris (proposition non(A)) ! C'est aussi simple que cela et c'est logiquement

imparable. Imaginer donc une situation dans laquelle on aurait A puis non(A) relève de l'absurdité logique, qui permet en retour d'affirmer n'importe quoi. Ainsi, en logique formelle le protocole CDR constitue une absurdité structurelle.

Certes on nous répondra que le cadre de cette logique formelle – celui donc associé à la rigueur du raisonnement – ne s'applique pas ici, car finalement l'événement A n'est vrai que pour une certaine période (on suppose ainsi qu'un médicament n'agit que pendant une certaine durée) et que longtemps après cette période on peut considérer que la proposition non(A) est vraie. Pourquoi pas ? Mais alors quel est le cadre formel dans lequel le protocole CDR constitue une « preuve » de l'imputabilité ? Dans le protocole retenu il y a manifestement des considérations temporelles (qui sont usuelles dans l'étude de la causalité) mais qui n'appartiennent cependant pas à la logique. De telles considérations nécessiteraient donc d'être explicitées et discutées de façon précise, c'est-à-dire dans un cadre mathématique. On attend toujours qu'il en soit ainsi. Dans l'idée de voir à quel point le discours de pharmacovigilance échappe à la formalisation, nous poursuivons le mail que nous a adressé le professeur spécialisé en pharmacovigilance, s'agissant de notre exploitation de la VAERS.

« Cependant, l'analyse VAERS que vous publiez est un non sens méthodologique (volontaire ou non) absolu. Ce que vous dites serait vrai (c'est du reste l'une de nos méthodes de détection des signaux) SI et SEULEMENT SI on était dans le cadre d'une surveillance épidémiologique dans laquelle tous les cas d'un événement donné sont recueillis (que ces cas aient été vaccinés ou non). Dans ce cas oui, on s'attend à une répartition aléatoire de la survenue des cas par rapport aux dates de vaccination. Dans le cas de VAERS, ceci ne tient pas puisque la base rassemble des cas que des personnes (professionnels de santé ou non) ont signalés parce qu'ils leurs semblaient possible/probable que l'événement puisse être lié au vaccin, essentiellement parce que le délai entre vaccination et événement leur semble évocateur. Les notifications sont en quelque sorte triées sur la base des délais d'apparition ce qui enlève tout sens à l'analyse que vous faites. Un peu comme si l'IFOP effectuait un sondage d'opinion dans la population d'un meeting de soutien à un candidat à la présidentielle. Il y a beaucoup de travaux l- dessus, il faut les respecter. Bien cordialement »

Pour recontextualiser la remarque de notre contradicteur (Bernard Bégaud), il faut rappeler en gros ce que nous affirmions lors de notre présentation qui exploitait des données de la VAERS : le fait que la proximité temporelle entre les décès et les injections constituait un indice sérieux de causalité entre la prise de la substance (la matériel génique tendant à faire produire la protéine Spike) et l'apparition de l'effet indésirable (la mort ou le symptôme) . Derrière cette question il y avait la volonté de répondre à la remarque suivante : il existe dans la VAERS (mais dans n'importe quelle base de données issue de la pharmacovigilance) des morts (ou des symptômes) qui succèdent effectivement à l'injection mais pour lesquelles cette injection n'est pas la cause du décès (ou de l'effet indésirable), de sorte que l'on doive en principe les écarter des statistiques de pharmacovigilance. Comme nous l'avions alors dit dans notre présentation, le fait que dans la VAERS la distribution des dates de mortalité (ou d'apparition des symptômes) ne soit pas uniforme sur les jours suivant l'injection indiquait qu'on avait là une forme de corrélation permettant de suspecter de façon claire et légitime une causalité entre les injections et les effets indésirables qui remontaient. En quoi cette remarque n'a pas plu à notre contradicteur ? Pour lui il semble évident que l'absence d'uniformité dans la variable aléatoire provient naturellement d'un biais d'échantillon et donc que notre étude « constitue un non-sens méthodologique ». Ce qu'il y a de frappant, là encore, dans la manière avec laquelle ce spécialiste de pharmacovigilance s'exprime, c'est l'absence de formalisation mathématique montrant qu'il maîtrise

sans doute assez mal l'utilisation des statistiques et des probabilités. Nous allons à nouveau rentrer dans des considérations mathématiques pour bien voir en quoi le langage de notre contradicteur relève assez peu de la science formalisée. Pour cela, il faut revenir à la définition du biais d'échantillon. A nouveau, dans ce qui va suivre, il va sans doute arriver que le lecteur perde le fil de la discussion dans la mesure où celle-ci s'engage dans le domaine de la formalisation mathématique. Mais alors il devra s'interroger lui-même sur la maîtrise qu'il possède ou non de la définition d'un biais d'échantillon. Et de la légitimité qu'il possède ou non – par ses discours – à entrer dans la discussion.

Définition 2 *Soit* Ω *un ensemble d'individus (par exemple l'ensemble des personnes vaccinées contre le covid en France). Soit* $C : \Omega \mapsto \{0,1\}$ *la variable aléatoire (i.e. la fonction) définie de la manière suivante : pour* ω *dans* Ω*, on a* $C(\omega) = 1$ *si l'individu* ω *est choisi pour l'échantillon, et* $C(\omega) = 0$ *dans le cas contraire.*

Définition 3 *Soit par ailleurs* $X : \Omega \mapsto \mathbb{R}$ *(l'ensemble des réels) une autre variable aléatoire (par exemple celle indiquant – pour chaque individu - le nombre de jour entre la date d'injection et la survenue d'un effet indésirable). Si p est une probabilité définie sur* $\mathscr{P}(\Omega)$*, on sait que cette variable aléatoire* X *va définir sur son ensemble image* $X(\Omega)$ *une probabilité* p_X*. C'est cette loi de probabilité que l'on essaye « d'estimer » à partir de l'échantillon.*

On voit ainsi qu'un biais d'échantillonage est défini par deux variables aléatoires :

1. Une variable aléatoire E permettant de décrire l'échantillon
2. Une variable aléatoire X permettant de décrire la loi que l'on cherche à établir

Pour comprendre comment on peut formaliser la question de l'échantillonnage et des biais qui lui sont imputables ou pas, on introduit l'ensemble $E \subset \Omega$ des personnes appartenant à l'échantillon (c'est-à-dire que l'on a ω dans E si et seulement si $C(\omega) = 1$). La fonction X peut ainsi être réduite au sous ensemble E.

On note par Y cette nouvelle variable aléatoire (qui ne peut-être notée de la même manière que X puisque son ensemble de définition a changé). De la même manière, on restreint la probabilité p à l'ensemble $\mathscr{P}(E)$ et on note par p_E la nouvelle probabilité ainsi obtenue (à nouveau on adopte une notation différente puisque l'ensemble de définition des probabilités a changé).

Définition 4 *Soit* $(\Omega, \mathscr{P}(\Omega), p)$ *un modèle de probabilité. Soit* $E \subset \Omega$ *tel que* $p(E) > 0$*. On appelle probabilité restreinte à E, noté par* p_E*, la probabilité définie sur* $\mathscr{P}(E)$ *telle que*

$$\forall B \subset E, \;\; p_E(B) = \frac{p(B)}{p(E)} \tag{3}$$

notons que la probabilité restreinte n'est pas exactement la probabilité conditionnelle dans la mesure où l'ensemble de définition ne correspond pas.

Définition 5 *On note alors par* p_Y *la loi de probabilité de la variable* Y *sur l'ensemble* $Y(E)$.

La définition d'un échantillonnage sans biais est alors la suivante

Définition 6 *on dit que l'échantillonnage E est sans biais lorsque*

1. $Y(E) = X(\Omega)$
2. *et que les lois de probabilités* p_X *et* p_Y *sont identiques.*

Evidemment on dira qu'un échantillon est biaisé si cette propriété n'est pas respectée. Cette définition peut être affaiblie si l'on impose seulement que les moyennes associées aux variables aléatoires X et Y soient identiques, ce que l'on retient usuellement de la définition d'un échantillon sans biais.

Nous allons reformuler, dans le cadre qui nous intéresse, la remarque énoncée par le professeur de pharmacovigilance et voir en quoi cette remarque manque de formalisation et devient par conséquent assez peu audible. Dans l'esprit du contradicteur, l'ensemble Ω n'est évidemment pas très bien défini (c'est une constante chez de nombreuses personnes essayant de faire des probabilités et des statistiques sans support mathématique). Il s'agit - pour reprendre les termes de ce dernier - de considérer *« le cadre d'une surveillance épidémiologique dans laquelle tous les cas d'un événement donné sont recueillis (que ces cas aient été vaccinés ou non) »*. Dit autrement, il semble que l'on doive construire l'ensemble Ω de la façon suivante : pour un effet indésirable donné (prenons la mort pour simplifier) on construit Ω comme étant l'ensemble des personnes décédées (disons sur la période couvrant la vaccination : ce point n'est évidemment pas précisé par l'auteur). Comme on le constate déjà, on est à la peine pour préciser quel est l'univers considéré.

Quelle est alors la variable aléatoire X que l'on doive définir sur cet ensemble Ω, de sorte que l'on puisse dire ensuite que la population de la VAERS en constitue un biais échantillonné ? On ne sait pas exactement. En fait, pour les personnes décédées dont les cas sont remontés via la VAERS, on sait parfaitement définir l'ensemble $E \subset \Omega$ et la variable Y associée (le temps entre la mort et la date d'injection). En revanche, eu égard à ce que nous propose le professeur comme définition de l'ensemble on ne sait pas du tout comment on doit définir la variable X. En fait on voit assez rapidement que l'évènement défini sur l'ensemble E (i.e les personnes vaccinées puis décédées et faisant remonter à la VAERS le nombre de jour entre l'injection et la mort) n'est pas susceptible d'être prolongé facilement sur l'ensemble que nus propose notre contradicteur. Donc, sauf à ce que l'auteur nous indique lui-même comment il faudrait procéder, il n'y a rien à comprendre de ce qu'il essaye de suggérer. C'est la raison pour laquelle cet éminent professeur a recours à une analogie avec les sondages politiques. Parce que dans cette situation, on est parfaitement capable de trouver précisément les ensembles et les variables aléatoires associées. Ainsi dan le cas du sondage politique on a :

Ω est l'ensemble des personnes qui votent. X est la variable aléatoire (binomiale) qui vaut 1 si on a l'intention de voter pour le candidat et qui vaut 0 si on n'a pas l'intention de voter pour lui. E est la fonction d'échantillon qui vaut 1 si l'individu (de Ω) est un participant du meeting et 0 sinon. La variable Y est donc maintenant celle qui décrit l'intention de vote *parmi* les individus présents au meeting.

Dans cette présentation, tout est merveilleusement défini du point de vue mathématique et constitue ainsi un modèle assez canonique de la notion de biais dans l'échantillonnage. Il semble ainsi totalement évident que les lois X et Y ne sont pas identiques et que les moyennes associées aux variables aléatoires X et Y ne le seront pas non plus.

Le problème n'est donc pas que l'exemple cité ne constitue pas un biais d'échantillonnage (celui de l'intention de vote), mais que le biais qui nous est reproché n'est pas défini par l'auteur de façon exploitable (celui du délai entre l'injection et la mort) et donc rende la remarque totalement caduque. De quoi exactement les données de la VAERS sont-elles un biais d'échantillon ? On ne le saura pas parce que l'auteur lui-même est incapable de le formuler mathématiquement la question qu'il pose autrement que sur un mode *préréflexif* et non scientifique qui emprunte seulement à l'analogie. Si bien qu'il est rigoureusement impossible de répondre quoi que ce soit de pertinent à sa remarque. Ce qui fait que la remarque de l'auteur n'a aucune

valeur, c'est l'idée que dans la définition du biais d'échantillon, toutes les grandeurs et les variables aléatoires sont déjà proprement définies sur l'ensemble , de sorte que la réduction à un sous ensemble E se construise de façon automatique et que la question du biais d'échantillon se pose de façon claire. Or dans son « raisonnement » l'auteur procède à l'inverse : il voit clairement qu'il y a une variable aléatoire défini sur le sous-ensemble E, mais il est incapable de « prolonger » cette variable aléatoire à un ensemble plus grand, de sorte que la question du biais se pose ensuite facilement.

En réalité, pour que les données de la VAERS puissent être replacées dans le cadre d'une discussion sur les biais d'échantillonnage, le mieux que l'on pourrait faire est sans doute la chose suivante :

L'ensemble Ω doit être constitué par l'ensemble des personnes vaccinées qui sont mortes ensuite. Pour que les choses soient un peu plus précises, on peut restreindre la définition de notre ensemble Ω de la façon suivante. On regarde dans la VAERS la plus grande durée séparant la date d'injection et la date de mort. Notons $Dmax$ (exprimée ne jours) cette quantité. On restreint ensuite l'ensemble comme étant celui de tous les individus vaccinés et décédés moins de $Dmax$ jours après leur injection (peu importe la raison putative pour laquelle ces individus sont décédés). On voit ainsi que les déclarations de la VAERS constituent bien un échantillon de Ω et que la variable X comptant pour un individu le nombre de jours entre son injection et sa mort est définissable sur l'ensemble Ω et sur sa restriction E (constitué par les cas remontés dans la VAERS). De cette manière la question du biais de l'échantillon est posée correctement.

Supposons donc que l'on se place dans le cadre précédent. Que voulait nous signifier notre contradicteur par sa phrase « Dans ce cas oui, on s'attend à une répartition aléatoire de la survenue des cas par rapport aux dates de vaccination » ? A nouveau le vocabulaire utilisé par l'auteur dénote une absence marquante de maîtrise du langage des statistiques et des probabilités. En cause le terme « aléatoire ». Que signifie en effet cet adjectif ? Du point de vue mathématique, absolument rien. D'ailleurs on peut largement affirmer que l'ensemble du travail des mathématiciens et des philosophes sur la question des probabilités et des statistiques avait précisément pour but de lever le flou et l'indétermination absolue entourant cette notion « d'aléatoire » pour la faire rentrer dans une maîtrise formelle et quantitative. Dit autrement la locution « distribution aléatoire » ne renvoie à rien. Pour tenter à nouveau de rectifier l'interpellation de notre contradicteur dans un cadre scientifique, il faut comprendre que la « répartition » qu'il évoque est celle associée à la loi X. Bien. Mais que signifie l'idée que cette distribution serait alors « aléatoire » ? Dans le vocabulaire des probabilités, la plupart des distributions (répartitions dans le vocable de l'interlocuteur) sont nommées de façon spécifique via l'introduction du terme de « loi » : loi gaussienne, loi de Poisson, loi binomiale, loi log-normale, loi uniforme etc. Il n'existe évidemment aucune loi qui porte le nom de « loi aléatoire ». Il est ainsi probable que dans l'esprit de notre interlocuteur la notion de « répartition aléatoire » désigne en réalité la locution « loi uniforme », mais on ne peut en être sûr tant ce professeur s'exprime de manière impropre s'agissant d'utiliser le vocabulaire adéquat du langage des probabilités. Finalement, ce que tentait de nous dire notre contradicteur était la chose suivante :

Sur l'ensemble $X(\Omega)$ la loi p_X associée à la variable aléatoire X est une loi uniforme, tandis que sur l'échantillon E des remontées de la VAERS, la loi p_Y n'est pas une loi uniforme.

Sauf qu'on ne voit pas en quoi cette remarque ferait de notre étude un *« non sens méthodologique »* : dans le problème proprement posé, l'ensemble Ω est celui des morts (toutes causes confondues) parmi les personnes vaccinées, tandis que l'échantillon E est celui des morts vaccinés pour lesquels on suppose que le vaccin est une cause probable de la mort. Dit autrement si l'on suppose (raisonnablement) que la loi p_X

est effectivement une loi uniforme (sur l'intervalle $[0, Dmax]$), alors il est nécessaire que la loi Y ne le soit pas puisqu'il existe précisément un biais sur les causes de la mort entre l'ensemble Ω et l'ensemble E. Par conséquent, l'idée qu'on observe une absence d'uniformité dans la loi p_Y et que cette loi représente un biais par rapport à la loi p_X (supposée uniforme) vient en réalité précisément renforcer l'idée que les personnes remontant dans la VAERS sont bien mortes du vaccin.

Il semble maintenant grand temps de conclure. Dans ces périodes troublées de politique sanitaire insensée, les faillites de la pharmacovigilance apparaissent au grand jour. L'insuffisance criante dans la formalisation des méthodes permet de rester (volontairement ou non) dans le flou, de sorte qu'il soit finalement impossible de faire émerger une vérité scientifique digne de ce nom. Qu'un ponte de la pharmacovigilance (Bernard Bégaud) affiche à ce point – sans même en avoir conscience – ses lacunes dans la modélisation scientifique de sa discipline en dit long sur l'absolue nécessité qu'il existe aujourd'hui de faire collaborer – dans un dialogue respectueux – les intervenants de différentes disciplines. Ecrire à notre propos que *« l'analyse VAERS que [nous publions] est un non sens méthodologique »* respecte assez peu le *« restons humbles et rigoureux »* que l'auteur revendique manifestement moins pour lui-même que pour les autres. La pharmacovigilance a essentiellement pour but d'alerter sur de « possibles » effets indésirables. Et c'est exactement ce que notre étude prétendait faire. En montrant que les chiffres bruts d'effets indésirables sur les injections immunogènes contre le COVID dépassaient en 9 mois ce qu'ils étaient par ailleurs en 30 ans sur l'ensemble des autres vaccins aux États-Unis, et en montrant – par le délai très court existant dans la VAERS entre les dates de morts et d'injection – que les mortalités enregistrées dans la VAERS démontraient vraisemblablement que les vaccins étaient bien la cause de la mort, notre intention était très claire : demander – au nom du *primum non nocere* – la fin de l'obligation vaccinale et inciter les pouvoirs publics à investiguer plus précisément dans la question de l'imputabilité. Car *in fine* la preuve ultime de causalité entre l'ingestion d'une substance et un effet indésirable qui lui serait imputable ne peut résulter que d'une étude physiopathologique ad hoc. Dans le cas des mortalités post vaccinales, il apparaît ainsi nécessaire de pratiquer systématiquement des autopsies qui permettraient par exemple de mesurer le taux de protéine Spike dans les tissus ou dans les fluides ayant conduit – par leur disfonctionnement – à entraîner la mort des personnes. De ce fait, on aurait des preuves de causalité qui échapperaient aux affirmations péremptoires des papes de la pharmacovigilance, qui ne résistent cependant pas très longtemps à une critique scientifique en règle.

.2 Code informatique de traitement et d'analyse pour le calcul de la mortalité post-vaccinale

Mortalite_post_vaccinale

April 5, 2022

Mortalité post vaccinale en Angleterre et au Portugal

Vincent PAVAN : mathématicien, auteur, enseignant-chercheur, Aix-Marseille Université

1 introduction

Dans ce document nous étudions la forme des courbes de mortalités dans certains pays ayant eu des épisodes de vaccinations. Nous nous intéressons à deux pays particuliers: l'Angleterre et le Portugal, dans des périodes qui suivent immédiatement le lancement des vaccinations, à l'entrée de l'hivers 2020. Pour mieux comprendre le choix que nous effectuons, nous allons d'abord commencer par observer les données OWID. NOus téléchargeons les données d'incidence de moratalité (i.e. le nombre quotidien de nouvelles personnes labelisées comme étant mortes du COVID) entre le 1er septebre 2020 et le 1er Juillet 2021.

1.1 téléchargement des données et régularisation

[101]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc ipyrmd
!pip install git+https://github.com/jupyter/nbconvert@master
from google.colab import output

!pip install -U -q PyDrive
from google.colab import drive
from google.colab import files
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

# Montage du drive et création des répertoires
drive.mount('/content/drive')
!mkdir /content/drive/MyDrive/MortalitePostVaccinale/
!mkdir /content/drive/MyDrive/MortalitePostVaccinale/datas/
output.clear()
```

[102]:
```
import pandas as pd
dfOWID = pd.read_csv("https://covid.ourworldindata.org/data/owid-covid-data.
↪csv")
```

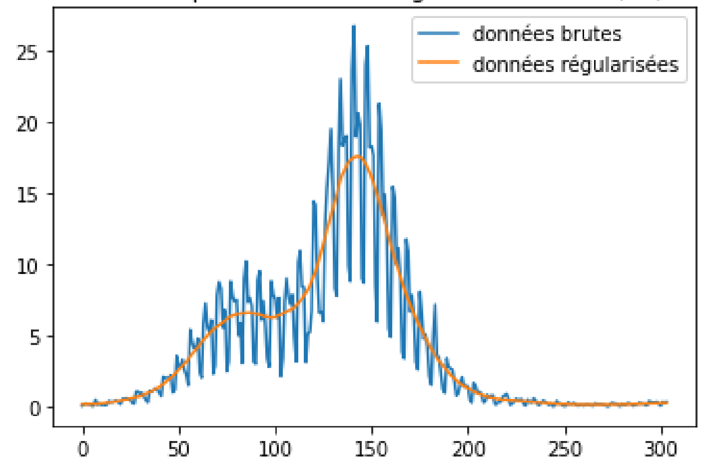
[103]:
```
list_variable = dfOWID.columns
#---------------------------------------------------------------
# sélectionne les données Portugaises et Anglaises et française
# entre le 1er septembre 2020 et le 1er Juillet 2021
#---------------------------------------------------------------
dfGBR = dfOWID [ (dfOWID.iso_code == "GBR") & (dfOWID.date >= "2020-09-01") &␣
↪(dfOWID.date <= "2021-07-01")]
dfPRT = dfOWID [ (dfOWID.iso_code == "PRT") & (dfOWID.date >= "2020-09-01") &␣
↪(dfOWID.date <= "2021-07-01")]
dfFRA = dfOWID [ (dfOWID.iso_code == "FRA") & (dfOWID.date >= "2020-09-01") &␣
↪(dfOWID.date <= "2021-07-01")]
PopGBR = dfGBR['population'].iloc[0]
PopPRT = dfPRT['population'].iloc[0]
PopFRA = dfFRA['population'].iloc[0]
```

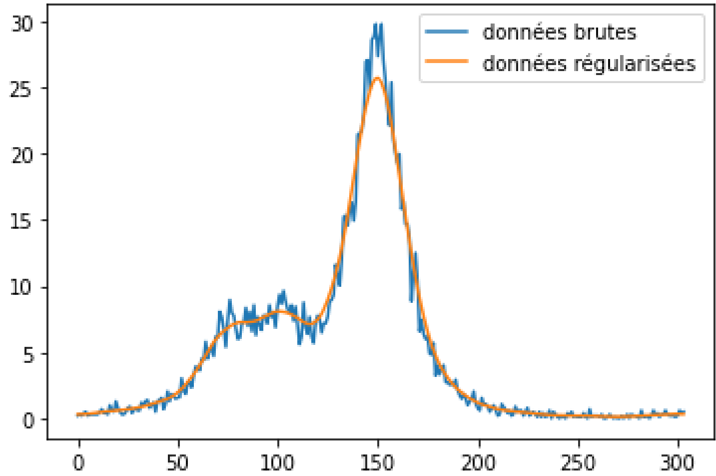
Ces données sont régularisées par une succession de deux moyennes glissantes centrées et son présentés en nombre quotidien de morts par million d'habitants.

[104]:
```
import numpy as np
def sym_moving_average(input_array, p):
  n = len(input_array)
  output_array = np.copy(input_array)
  if (n < 2*p + 1):
    return output_array
  else:
    for k in range(0,p):
      output_array[k] = 2/(2*p+1)*np.sum(input_array[0:p-k]) + 1/(2*p+1)*np.
↪sum(input_array[p-k:p+k+1])
    for k in range(p,n-p):
      output_array[k] = 1/(2*p+1) * np.sum(input_array[k-p:k+p+1])
    for k in range(n-p,n):
      output_array[k] = 2/(2*p+1)*np.sum(input_array[2*n-p-k-1:n]) + 1/
↪(2*p+1)*np.sum(input_array[k-p:2*n-p-k-1])
    return output_array
```

[105]:
```
import matplotlib.pyplot as plt
new_death_GBR = dfGBR['new_deaths'].values
smooth_new_death_GBR = sym_moving_average(new_death_GBR, 8)
smooth_new_death_GBR = sym_moving_average(smooth_new_death_GBR, 3)
new_death_PRT = dfPRT['new_deaths'].values
smooth_new_death_PRT = sym_moving_average(new_death_PRT, 8)
smooth_new_death_PRT = sym_moving_average(smooth_new_death_PRT, 3)
plt.clf()
```

```
plt.plot(new_death_GBR/PopGBR*1e6, label = "données brutes")
plt.plot(smooth_new_death_GBR/PopGBR*1e6, label = "données régularisées")
plt.legend(loc = "upper right")
plt.title("incidence de mortalité (morts par million) de l'Angleterre entre le␣
 ↪1/09/2020 et le 1/07/2021")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_incidence_GBR.
 ↪png")
plt.show()
plt.clf()
plt.plot(new_death_PRT/PopPRT*1e6, label = "données brutes")
plt.plot(smooth_new_death_PRT/PopPRT*1e6, label = "données régularisées")
plt.legend(loc = "upper right")
plt.title("incidence de mortalité (morts par million) du Portugal entre le 1/09/
 ↪2020 et le 1/07/2021")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_incidence_PRT.
 ↪png")
plt.show()
```





1.2 Observation caractéristique

Après observation sur le site OurWorlInData, on décide qu'il existe pour les deux pays un "point anguleux" (c'est à dire une forme de discontinuité dans la valeur de la dérivée):

1. Pour le Royaume-Unis: le 28 décembre 2020 (c'est à dire 118 jours après le 1er septembre 2020)
2. Pour le Portugal: le 4 Janvier 2020 (c'est à dire 125 jours après le 1er septembre 2020)

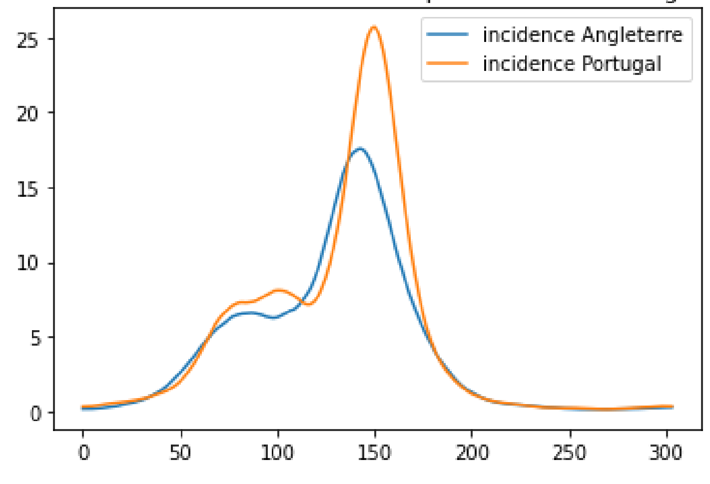
[106]:
```
import datetime
#------------------------------------------
# repère des dates importantes
#-----------------------------------------
# début d'observation des nouvelles émidémies
#---------------------------------------------
dt0 = datetime.datetime(2020, 9, 1)
#-----------------------------------------
# date des points anguleux
#-----------------------------------------
dtGBR =  datetime.datetime(2020, 12, 28)
dtPRT =  datetime.datetime(2021, 1, 4)
#-------------------------------------------
# date officielle de début de vaccination
#-------------------------------------------
dtVaccGBR = datetime.datetime(2020,12,8)
dtVaccPRT = datetime.datetime(2020,12,27)
#-----------------------------------------------
print("date de début de vaccination au Royaume-Unis:", dtVaccGBR.date())
print("date de point anguleux au Royaume-Unis:", dtGBR.date())
```

```
print("nombre de jours écoulés depuis le", dt0.date(), "pour le point anguleux␣
 ↪du Royaume Unis:", (dtGBR - dt0).days + 1, " jours")
print("date de début de vaccination au Portugal:", dtVaccPRT.date() )
print("date de point anguleux au Royaume-Unis:", dtGBR.date())
print("nombre de jours écoulés depuis le", dt0.date(), "pour le point anguleux␣
 ↪du Portugal:", (dtPRT - dt0).days + 1, "jours")
```

```
date de début de vaccination au Royaume-Unis: 2020-12-08
date de point anguleux au Royaume-Unis: 2020-12-28
nombre de jours écoulés depuis le 2020-09-01 pour le point anguleux du Royaume
Unis: 119  jours
date de début de vaccination au Portugal: 2020-12-27
date de point anguleux au Royaume-Unis: 2020-12-28
nombre de jours écoulés depuis le 2020-09-01 pour le point anguleux du Portugal:
126 jours
```

Modulo ces dates, on voit que les deux pays suivent deux profils de courbe qui se ressemblent effectivement de façon interpelante. Une forme de synchronicité apparaît dans les premiers temps: progression d'une épidémie où le nombre de mort quotidien par million d'habitant est identique. Puis un léger décalage se produit et finalement les profils se rejoignent vers la fin du moi de février.

[107]:
```
plt.clf()
plt.plot(smooth_new_death_GBR/PopGBR*1e6, label = "incidence Angleterre")
plt.plot(smooth_new_death_PRT/PopPRT*1e6, label = "incidence Portugal")
plt.legend(loc = "upper right")
plt.title("comparaison des incidences de mortalité ((morts par million) du␣
 ↪Portugal et de l'Angleterre")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_comp_incidence.
 ↪png")
plt.show()
```



On peut ”mesurer” le décalage survenant lors du déphasage des courbes en estimant la différence des dates de pic. On obtient alors

[108]:
```
maxGBR = np.where(smooth_new_death_GBR == np.max(smooth_new_death_GBR))[0][0]
maxPRT = np.where(smooth_new_death_PRT == np.max(smooth_new_death_PRT))[0][0]
print("le déphase des maxima entre le Royaume Uni et le Protugal est de:",␣
 ↪maxPRT - maxGBR, " jours")
```

```
le déphase des maxima entre le Royaume Uni et le Protugal est de: 7  jours
```

Ces changements de ”régime” épidémique sont déterminés de façon observationnelle (c'est à dire sans algorithme mathématiques particulier). Ils semblent indiquer - dans le processus de propagation - de l'irruption d'un phénomène qui aurait eu tendance à ”redistribuer” et à forcer l'évolution selon certaines caractéristiques. Qualitativement ces caractéristiques semblent être les suivantes:

1. Une symétrie dans la forme de la courbe
2. Une cinétique rapide

Ces deux aspects semblent s'écarter des caractérisitques usuelles des courbes de mortalités épidémiques. Fondamentalement, dans les modèles SIR, les courbes sont asymétriques (l'asymétrie étant d'ailleurs caractéristique du rapport $\mathcal{R}_0 = \beta/\gamma$). D'autre part la cinétique est en général plus lente que celle qui apparaît. Il aparaît immanquablement que ces changements brusques dans la courbe épidémiologique sont incluses dans la période vaccinale.

1.3 Question concernant les courbes de mortalité du Prtugal et de l'Angleterre

La question qui se pose est dnc de savoir si l'on peut caractériser deux types d'évènements épidémiologiques:

1. Un évènement de type diffusion ”classique” qui serait modélisable par une courbe de type SIR
2. Un évènement de type propagation ”non classique” qui serait modélisable avec une courbe de forme différente à celle des SIR

2 Rechercher des Maxweliennes

2.1 Présentation

Les Maxwelliennes sont des courbes importantes des mathématiques. Elles interviennent notoirement de façon prépondérantes en probabilité grâce au théorème de la ”Limite Centrale”. Les courbes Maxwelliennes modélisent ainsi de nombreux phénomènes aléatoires. Rappelons mathématiquement ce qu'est une Maxwellienne. Il s'agit d'une fonction définie par:

$$\forall t \in \mathbb{R}, \quad \mathcal{M}(t) = \frac{n}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right) \tag{1}$$

Dans cette écriture, t_m est la valeur moyenne, tandis que σ désigne l'écart et n la densité. Ces valeurs sont les paramètres de la Maxwellienne. Ces paramètres sont liés à la fonction grâce aux intégrales suivantes:

$$n = \int_{-\infty}^{\infty} \mathcal{M}(t)\, dt \tag{2}$$

$$nt_m = \int_{-\infty}^{\infty} t\mathcal{M}(t)\, dt \tag{3}$$

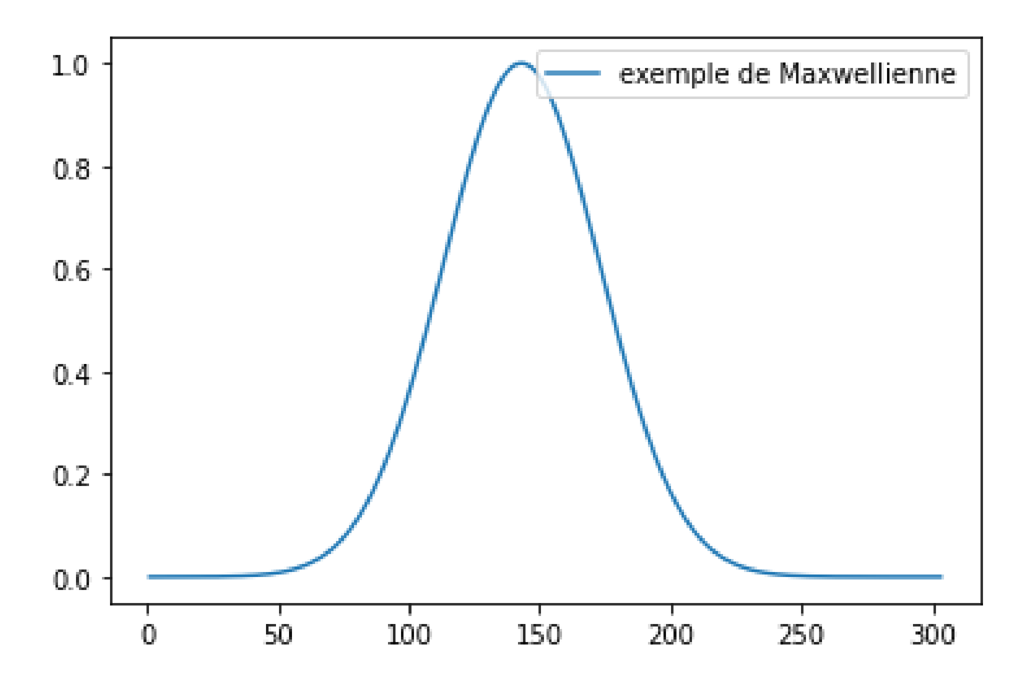
$$n\left(\sigma^2 + t_m^2\right) = \int_{-\infty}^{\infty} t^2\mathcal{M}(t)\, dt \tag{4}$$

La question de l'identification des Maxwelliennes est une question importante des problèmes inverses. Comment savoir si une courbe est une Maxwellienne ou non ? *Plus difficile: comment savoir si une partie de courbe est une Maxwellienne ?* Il se trouve en effet que l'idée qui nous anime c'est de montrer qu'une partie des courbes de mortalité de l'Angleterre et du Portugal, sur la période de l'hiver 2020-2021 peuvent effectivement être modélisées par des Maxwelliennes. Ce qui nous permettra de séparer, dans d'autre période, ce qui ne viendrait pas de cet aspect Maxwellien. Une Maxwellienne est donc caractérisée par trois paramètres:

1. Sa densité n
2. sa moyenne t_m
3. son écart type σ

On montre ci-dessous un exemple de courbe Maxellienne:

```
[109]: t = np.arange(1,len(smooth_new_death_GBR))
t_m = maxGBR
s = 30 # jour^2
plt.clf()
plt.plot(t, np.exp(-0.5*(t-t_m)**2/s**2), label  = "exemple de Maxwellienne")
plt.legend(loc = "upper right")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_Maxwellienne.
 ↪png")
plt.show()
```



On appelle Maxwellienne discrète toute suite $(\xi_i : i \in [1, N])$ de la forme:

$$\forall i \in [1, N], \quad \xi_i = \frac{n}{Z(m, s)} \exp\left(-\frac{(x_i - m)^2}{2s^2}\right) \tag{5}$$

$$Z(m, s) = \sum_{i=1}^{i=N} \exp\left(-\frac{(x_i - m)^2}{2s^2}\right) \tag{6}$$

Au besoin, pour spécifier la dépendance de $\boldsymbol{\xi}$ avec ses paramètres, on notera $\boldsymbol{\xi}\left[n, m, s^2\right]$

[110]:
```
def momentmap(f,t, meas):
  d = np.sum(f * meas)
  q = np.sum(f * t * meas)
  e = np.sum(f* t**2 * meas)
  return np.array([d,q,e])


def paramexp(param, t):
  #-------------------------------------------------
  # calcule la valeur d'une Maxwelienne paramétrée
  #-------------------------------------------------
  n = param[0]
  m = param[1]
  s = param[2]
  Zms = np.sum(np.exp(-(t - m)**2/2/s))
```

```
  return n/Zms * np.exp(-(t - m)**2/2/s)


def momparamexp(param, t, meas):
  #----------------------------------------
  # calcule les moments d'une maxwellienne
  # paramétrée
  #---------------------------------------
  f = paramexp(param, t)
  mom = momentmap(f,t, meas)
  return mom
```

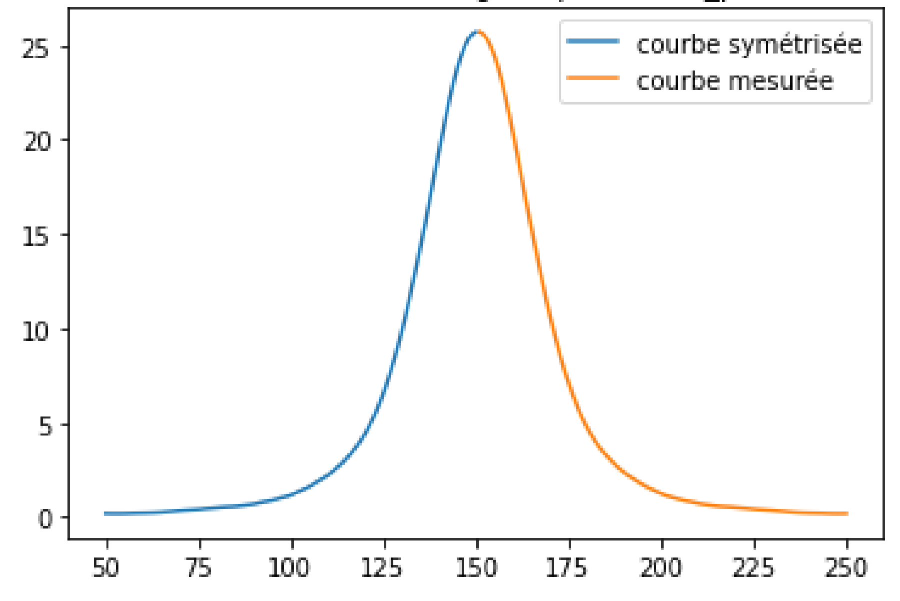
2.2 Exploration préparatoire

Dans l'étude que nous proposons, tout ne peut pas être rendu totalement algorithmique: la courbe que nous étudions est complexe. En particulier on peine par exemple à trouver un critère mathématique satisfaisant pour décider du "point anguleux" (en réalité il s'agit d'une croissance rapide de la dérivée, sans que l'on sache trop comment traduire ce "rapide" et comment décider du moment de retournement). De la même manière, on peine à l'idée de savoir quel intervalle de temps est celui pour lequel on va commencer à établir l'existence d'un régime Maxwellien. Il faut donc commencer par "tatonner" avant de rendre les choses plus algorithmiques.

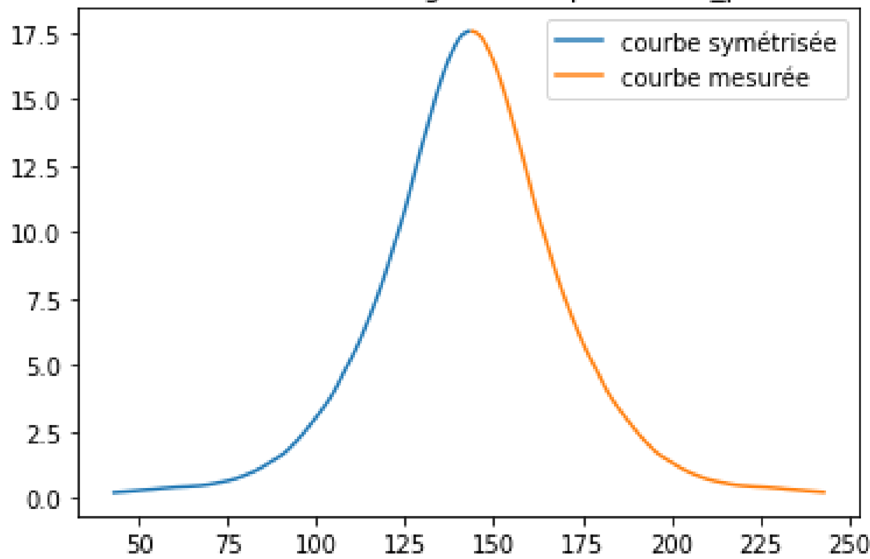
D'un point de vue graphique, il semble raisonnable de penser que dans les courbes de mortalité, les évènements qui se situent après les pics épidémiques appartiennent à une unité de diffusion: autrement dit après le pic, on considère qu'il n'y a plus qu'une sorte de phénomène épidémiologique et non pas la superposition de plusieurs phénomènes. C'est ici, dans cet intervalle que l'on va considérer de façon "objective" que la courbe est raisonnablement représentée par une Maxwelienne. Pour voir si cette première approximation est possible, on va d'abord symétriser les données de $t \geq t_p$ sur les données $t \leq t_p$ où t_p représente le temps de pic, c'est à dire en fait le maximum de la courbe. ON choisit donc Δ un demi intervalle et on va donc symétriser la courbe des données sur l'intervalle $[t_m - \Delta, t_m + \Delta]$ en partant des données d'incidence sur l'intervalle $[t_m, t_m + \Delta]$.

[111]:
```
#--------------------------------------------
# on prépare la symétrisation pour le calcul
# des maxwelliennes
#--------------------------------------------
Delta = 100
tGBR_sym = np.arange(maxGBR - Delta, maxGBR + Delta+1, 1)
tPRT_sym = np.arange(maxPRT - Delta, maxPRT + Delta+1, 1)
sym_GBR = np.zeros(len(smooth_new_death_GBR))
sym_PRT = np.zeros(len(smooth_new_death_PRT))
for i in range(0, Delta+1):
  sym_GBR[maxGBR-i] = smooth_new_death_GBR[maxGBR+i]/PopGBR*1e6
  sym_GBR[maxGBR+i] = smooth_new_death_GBR[maxGBR+i]/PopGBR*1e6
  sym_PRT[maxPRT-i] = smooth_new_death_PRT[maxPRT+i]/PopPRT*1e6
  sym_PRT[maxPRT+i] = smooth_new_death_PRT[maxPRT+i]/PopPRT*1e6
```

```
plt.clf()
plt.plot(tPRT_sym[0:Delta+1], sym_PRT[(maxPRT - Delta):maxPRT+1], label =␣
↪"courbe symétrisée")
plt.plot(tPRT_sym[Delta+1:2*Delta+1], sym_PRT[maxPRT:(maxPRT+Delta)], label =␣
↪"courbe mesurée")
plt.legend(loc = "upper right")
plt.title("symétrisation des données du Portugal à partir de t_p = 150 et D =␣
↪100")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_symetriePRT.png")
plt.show()
plt.clf()
plt.plot(tGBR_sym[0:Delta+1], sym_GBR[(maxGBR - Delta):maxGBR+1], label =␣
↪"courbe symétrisée")
plt.plot(tGBR_sym[Delta+1:2*Delta+1], sym_GBR[maxGBR:(maxGBR+Delta)], label =␣
↪"courbe mesurée")
plt.legend(loc = "upper right")
plt.title("symétrisation des données de l'Angleterre à partir de t_p = 143 et D␣
↪= 100")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_symetrieGBR.png")
plt.show()
```







A partir de ces courbes, si l'on suppose que l'on est Maxwellien, on peut alors trouver approximativement les valeurs de n, t_m et σ grâce aux calculs des moments via les intégrales dédiées. En utilisant la discrétisation temporelle on peut donc avoir une approximation numérique de ce résultats

Plus précisément, si $(s_i : i \in [-P, P])$ est l'ensemble des valeurs de la courbe symétrisée, et si V_i désigne la mesure de chaque cellule dont est extrait la valeur, on pose

$$n = \sum_{i=-P}^{i=P} s_i V_i \tag{7}$$

$$t_p \approx \frac{1}{n} \sum_{i=-P}^{i=P} t_i s_i V_i \tag{8}$$

$$\sigma \approx \left(\frac{1}{n} \sum_{i=-P}^{i=P} t_i^2 s_i V_i \right) - t_p^2 \tag{9}$$

[112]:
```
meas = 1 #jour
momPRT_sym = momentmap(sym_PRT[tPRT_sym], tPRT_sym, meas)
momGBR_sym = momentmap(sym_GBR[tGBR_sym], tGBR_sym, meas)
muGBR_sym = momGBR_sym[1]/momGBR_sym[0]
muPRT_sym = momPRT_sym[1]/momPRT_sym[0]
sGBR_sym = 1/momGBR_sym[0] * momGBR_sym[2] - muGBR_sym**2
sPRT_sym = 1/momPRT_sym[0] * momPRT_sym[2] - muPRT_sym**2
```

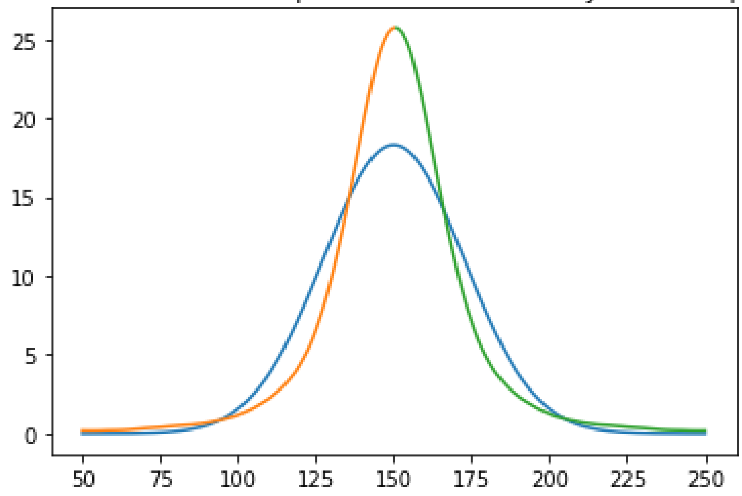
```
print("Pour le Royaume Uni le moment symétrique est de", momGBR_sym[0], "morts␣
↪par million d'habitant")
print("Pour le Royaume Uni le temps de pic est", muGBR_sym, "jours")
print("Pour le Royaume Uni l'écart type maxwellien", np.sqrt(sGBR_sym),␣
↪"jours^2")
print("Pour le Portugal le moment symétrique est de", momPRT_sym[0], "morts par␣
↪million d'habitant")
print("Pour le Portugal le temps de pic est", muPRT_sym, "jours")
print("Pour Portugal l'écart type maxwellien", np.sqrt(sPRT_sym), "jours^2")
```

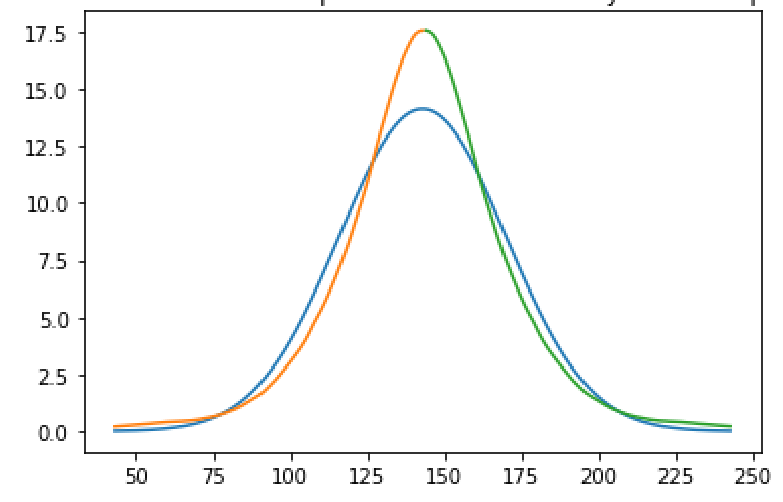
```
Pour le Royaume Uni le moment symétrique est de 955.6361653705151 morts par
million d'habitant
Pour le Royaume Uni le temps de pic est 143.0 jours
Pour le Royaume Uni l'écart type maxwellien 27.002688717633777 jours^2
Pour le Portugal le moment symétrique est de 1034.9783168040092 morts par
million d'habitant
Pour le Portugal le temps de pic est 150.0 jours
Pour Portugal l'écart type maxwellien 22.55741697514642 jours^2
```

Comme les courbes ne peuvent pas être exactement des Maxwelliennes sur l'ensemble des valeurs considérées de t, l'écart typeque l'on calcule ne peut pas être le bon. Si l'on trace ainsi les Maxwelliennes d'après les paramètres que l'on vient de calculer on trouver alors:

[113]:
```
momPRT = momentmap(sym_PRT[tPRT_sym], tPRT_sym, meas)
momGBR = momentmap(sym_GBR[tGBR_sym], tGBR_sym, meas)
muGBR = momGBR[1]/momGBR[0]
muPRT = momPRT[1]/momPRT[0]
sGBR = 1/momGBR[0] * momGBR[2] - muGBR**2
sPRT = 1/momPRT[0] * momPRT[2] - muPRT**2
param_GBR = np.array([momGBR[0], muGBR, sGBR])
param_PRT = np.array([momPRT[0], muPRT, sPRT])
plt.clf()
plt.plot(tPRT_sym, paramexp(param_PRT, tPRT_sym))
plt.plot(tPRT_sym[0:Delta+1], sym_PRT[(maxPRT - Delta):maxPRT+1])
plt.plot(tPRT_sym[Delta+1:2*Delta+1], sym_PRT[maxPRT:(maxPRT+Delta)], label =␣
↪"Maxwellienne brute")
plt.title("Maxwellienne de même moments que ceux des données symétrisées pour␣
↪le Portugal")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_MAxbrutePRT.png")
plt.show()
plt.clf()
plt.plot(tGBR_sym, paramexp(param_GBR, tGBR_sym))
plt.plot(tGBR_sym[0:Delta+1], sym_GBR[(maxGBR - Delta):maxGBR+1])
plt.plot(tGBR_sym[Delta+1:2*Delta+1], sym_GBR[maxGBR:(maxGBR+Delta)], label =␣
↪"Maxwellienne brute")
plt.title("Maxwellienne de même moments que ceux des données symétrisées pour␣
↪l'Angleterre")
```

```
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_MAxbruteGBR.png")
plt.show()
```





2.3 Maxwelliennes optimales

L'accord n'est pas suffisant. Pour obtenir un meilleur accord, on va jouer sur le paramètre σ. En effet, le maximum doit être obtenu pour $t = t_p$ et le moment d'ordre un (en l'occurrence le nombre n total de décès par million d'habitant) doit être préservé. On va donc trouver s qui minimize le problème

$$\inf_{\sigma\in\mathbb{R}^+}\left(\sup_{i\in[1,N]}|s_i-\xi_i[n,t_p,\sigma]|\right) \tag{10}$$

[114]:
```
import scipy
from scipy import misc
from numpy import linalg as la
```

[115]:
```
def Ldiff(f, s, param, t, meas, deg):
  paramloc = np.copy(param)
  paramloc[2] = s
  err = la.norm(f - paramexp(paramloc, t), deg)
  return err
```

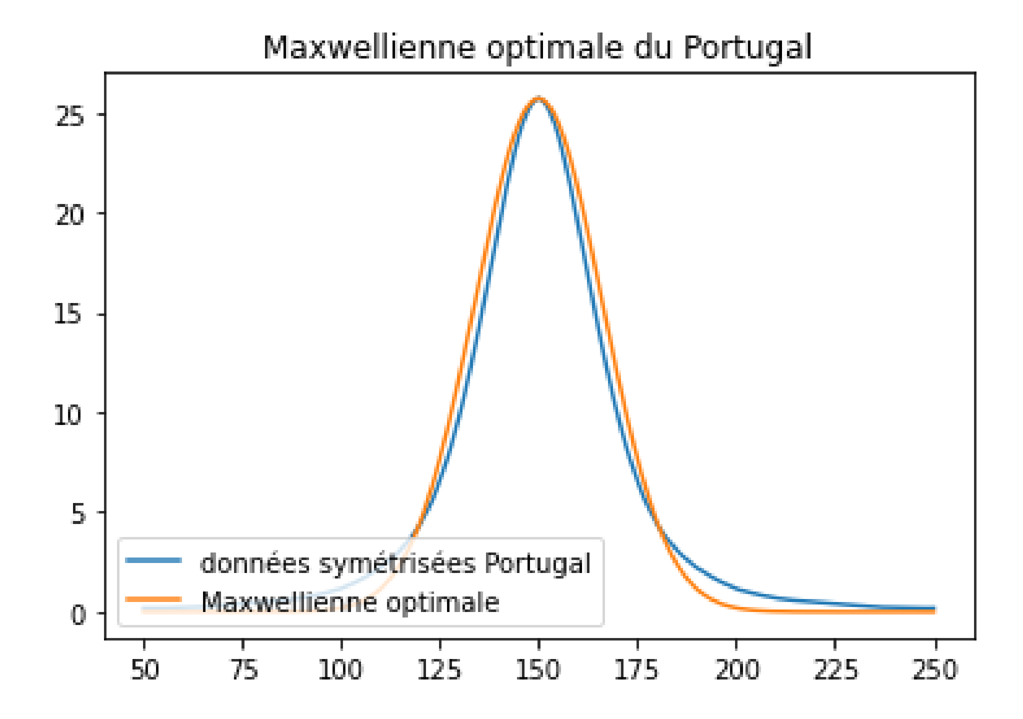
On peut appliquer un algorithme "exhaustif" qui permet de trouver ssez rapidement les paramètres (discrets) de la Maxwellienne optimale

[116]:
```
r = np.infty
P = 10001
tPRT_opt = np.arange(maxPRT-30,maxPRT+31,1)
deg = np.inf
for i in range(1,P):
  r1 = Ldiff(smooth_new_death_PRT[tPRT_opt]/PopPRT*1e6, i/P*param_PRT[2] ,␣
 ↪param_PRT, tPRT_opt, meas, deg)
  if (r1 <= r):
    r = r1
    s = i/P*param_PRT[2]

sPRT = s
param_PRT[2] = sPRT
print("écart type optimal", np.round(np.sqrt(s),2), "jours")
err_rel_PRT = la.norm(sym_PRT[tPRT_sym] - paramexp(param_PRT,tPRT_sym),np.inf)/
 ↪la.norm(sym_PRT[tPRT_sym],np.inf)
print("erreur relative:", np.round(err_rel_PRT,4), "en norme", np.inf )
plt.clf()
plt.plot(tPRT_sym, sym_PRT[tPRT_sym], label = "données symétrisées Portugal")
plt.plot(tPRT_sym,paramexp(param_PRT, tPRT_sym), label = "Maxwellienne␣
 ↪optimale")
plt.legend(loc = "lower left")
plt.title("Maxwellienne optimale du Portugal")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_MAxoptPRT.png")
plt.show()
d95 = datetime.timedelta(int(maxPRT - 1.960*np.sqrt(param_PRT[2])))
d99 = datetime.timedelta(int(maxPRT - 2.576*np.sqrt(param_PRT[2])))
print(d95)
dateinit95 = dt0+d95
dateinit99 = dt0+d99
```

```
print(dateinit99, dateinit95)
```

```
écart type optimal 16.07 jours
erreur relative: 0.0899 en norme inf
```




```
118 days, 0:00:00
2020-12-18 00:00:00 2020-12-28 00:00:00
```

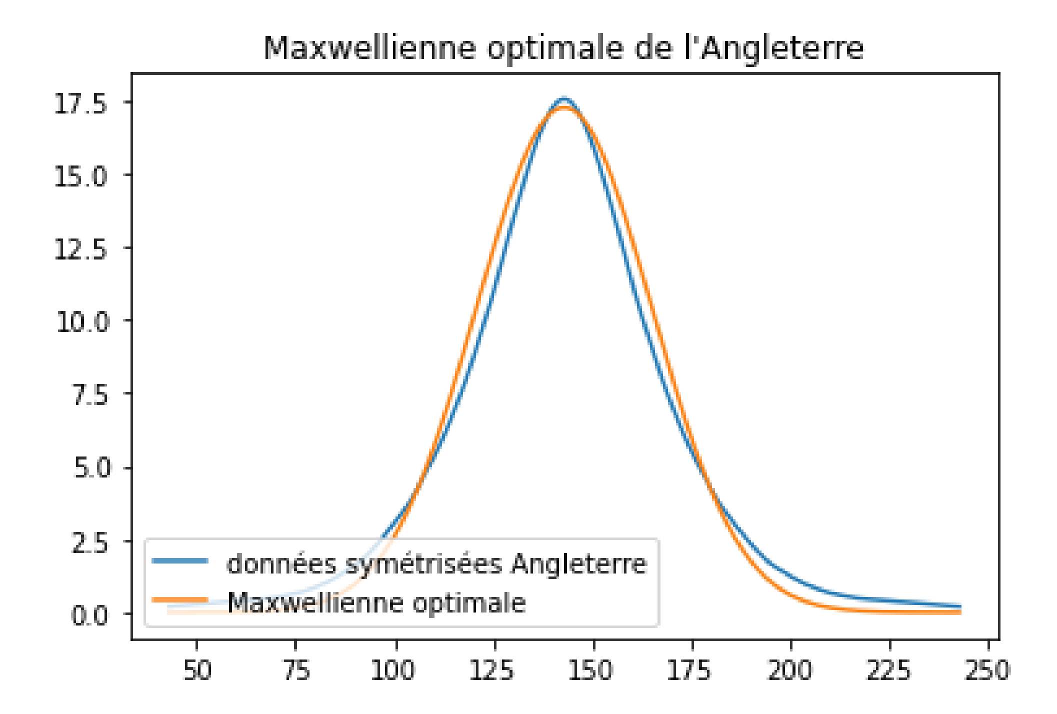
[117]:
```
r = np.infty
P = 10001
tGBR_opt = np.arange(maxGBR-30,maxGBR+31,1)
for i in range(1,P):
  r1 = Ldiff(smooth_new_death_GBR[tGBR_opt]/PopGBR*1e6, i/P*param_GBR[2] ,␣
↪param_GBR, tGBR_opt, meas, np.inf)
  if (r1 <= r):
    r = r1
    s = i/P*param_GBR[2]

sGBR = s
param_GBR[2] = sGBR
err_rel_GBR = la.norm(sym_GBR[tGBR_sym] - paramexp(param_GBR,tGBR_sym),np.inf)/
↪la.norm(sym_GBR[tGBR_sym],np.inf)
print("erreur relative:", np.round(err_rel_GBR,4), "en norme", np.inf )

plt.clf()
```

```
plt.plot(tGBR_sym, sym_GBR[tGBR_sym], label = "données symétrisées Angleterre")
plt.plot(tGBR_sym,paramexp(param_GBR, tGBR_sym), label = "Maxwellienne␣
 ↪optimale")
plt.legend(loc = "lower left")
plt.title("Maxwellienne optimale de l'Angleterre")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_MAxoptGBR.png")
plt.show()
```

```
erreur relative: 0.086 en norme inf
```



3 séparation du signal en composantes

3.1 Visualisation des processus Maxwelliens

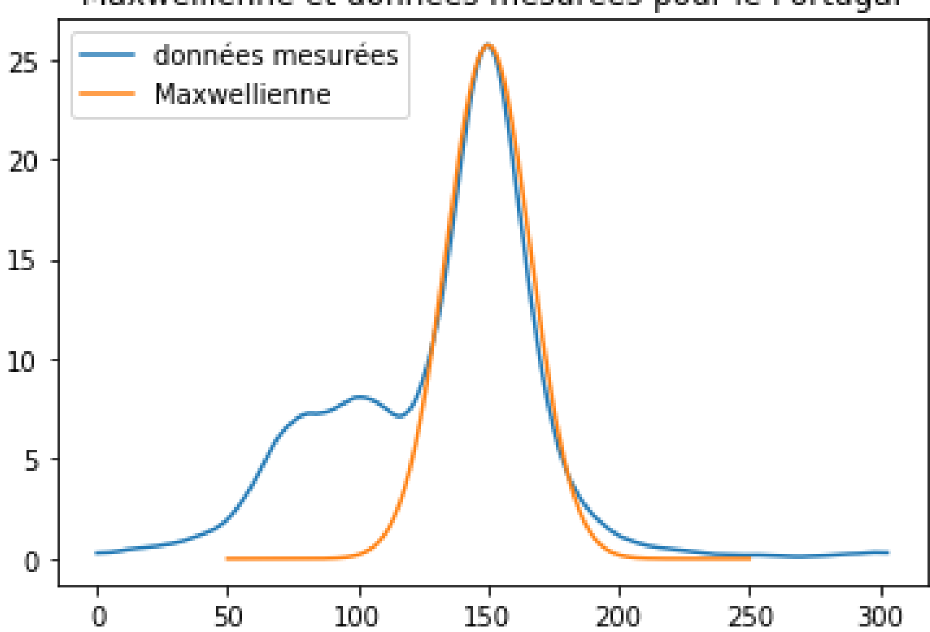
On peut maintenant superposer les processus Maxwelliens aux courbes de mortalité. On obtient ainsi:

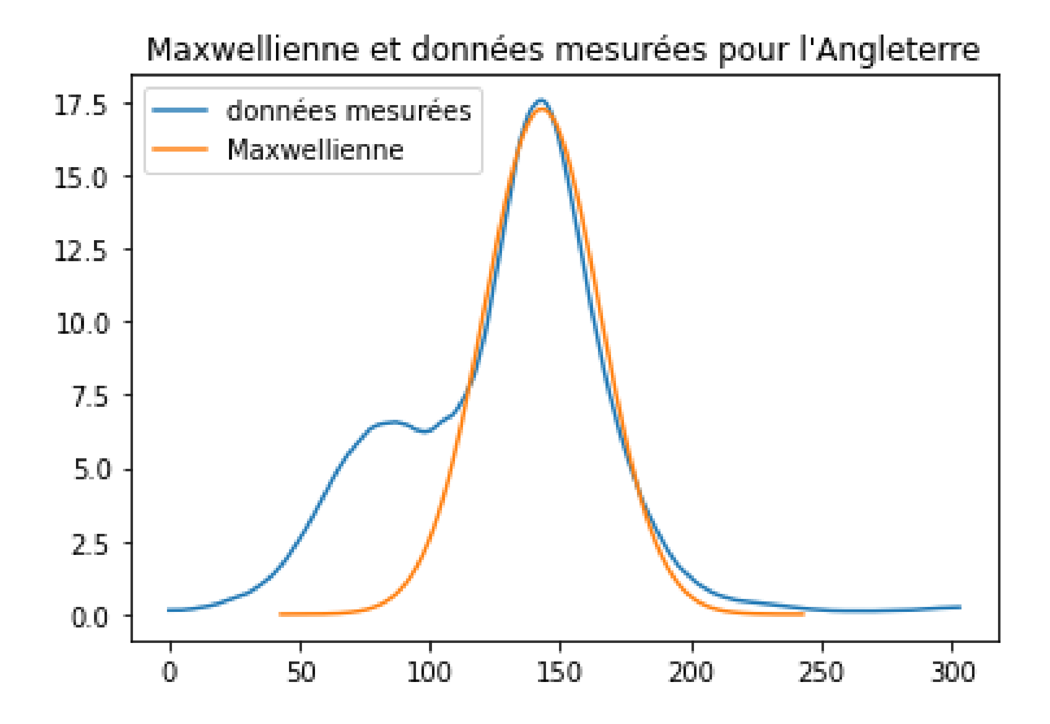
[118]:
```
plt.clf()
plt.plot(smooth_new_death_PRT/PopPRT*1e6, label = "données mesurées")
plt.plot(tPRT_sym, paramexp(param_PRT, tPRT_sym), label = "Maxwellienne")
plt.legend(loc = "upper left")
plt.title("Maxwellienne et données mesurées pour le Portugal")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_reelMaxPRT.png")
```

```
plt.show()

plt.clf()
plt.plot(smooth_new_death_GBR/PopGBR*1e6, label = "données mesurées")
plt.plot(tGBR_sym, paramexp(param_GBR, tGBR_sym), label = "Maxwellienne")
plt.legend(loc = "upper left")
plt.title("Maxwellienne et données mesurées pour l'Angleterre")
plt.savefig("/content/drive/MyDrive/MortalitePostVaccinale/fig_reelMaxGBR.png")
plt.show()
```







Comme on peut le constater, il existe effectivement une période significative où l'aproximation Maxwellienne est très pertinente. Nous allons affiner les choses de la façon suivante:

1. On va sélectionner la plage où la correspondance Maxwellienne est pertinente
2. Sur cette plage on va automatiser la recherche des paramètres Maxwelliens de la courbe.

3.2 Origine du phénomène Maxwellien

On peut maintenant aller rechercher l'origine du phénomène Maxwellien. Bien que la courbe soit définie a priori sur $\mathbb{R}$ entier, on peut définir l'origine du pénomène Maxwellien comme étant quelque part entre

1. La bonre inférieure de l'intervalle de confiance à 99%
2. La borne inférieure de l'intervalle de confiance à 95%

[119]:
```
d95GBR = datetime.timedelta(int(maxGBR - 1.960*np.sqrt(param_GBR[2])))
d99GBR = datetime.timedelta(int(maxGBR - 2.576*np.sqrt(param_GBR[2])))
dateinit95GBR = dt0+d95GBR
dateinit99GBR = dt0+d99GBR
print("Pour le Royaume Unis, début du régime Maxwellien entre le␣
 ↪",dateinit99GBR.date(), "et le", dateinit95GBR.date())
print("date de début de vaccination au Royaume-Unis:", dtVaccGBR.date() )

d95PRT = datetime.timedelta(int(maxPRT - 1.960*np.sqrt(param_PRT[2])))
d99PRT = datetime.timedelta(int(maxPRT - 2.576*np.sqrt(param_PRT[2])))
```

```
dateinit95PRT = dt0+d95PRT
dateinit99PRT = dt0+d99PRT
print("Pour le Portugal, début du régime Maxwellien entre le ", dateinit99PRT.
 ↪date(), "et le ", dateinit95PRT.date())
print("date de début de vaccination au Portugal:", dtVaccPRT.date())
```

```
Pour le Royaume Unis, début du régime Maxwellien entre le  2020-11-26 et le
2020-12-09
date de début de vaccination au Royaume-Unis: 2020-12-08
Pour le Portugal, début du régime Maxwellien entre le  2020-12-18 et le
2020-12-28
date de début de vaccination au Portugal: 2020-12-27
```

3.3 production des données et des rapports

[ ]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc ipyrmd
!pip install git+https://github.com/jupyter/nbconvert@master
from google.colab import output

!pip install -U -q PyDrive
from google.colab import drive
from google.colab import files
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

# Montage du drive et création des répertoires
drive.mount('/content/drive')
!mkdir /content/drive/MyDrive/MortalitePostVaccinale/
!mkdir /content/drive/MyDrive/MortalitePostVaccinale/datas/
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

downloaded = drive.CreateFile({"id":"1VKLd3enKNwPhQqI_q50kgVRGmEy2D1dH"})
downloaded.GetContentFile("/content/drive/MyDrive/MortalitePostVaccinale/
 ↪Mortalite_post_vaccinale.ipynb")
!jupyter nbconvert "/content/drive/MyDrive/MortalitePostVaccinale/
 ↪Mortalite_post_vaccinale.ipynb" --to pdf
output.clear()
```

```
Reading package lists… Done
```

```
Building dependency tree
Reading state information… Done
pandoc is already the newest version (1.19.2.4~dfsg-1build4).
texlive is already the newest version (2017.20180305-1).
texlive-latex-extra is already the newest version (2017.20180305-2).
texlive-xetex is already the newest version (2017.20180305-1).
0 upgraded, 0 newly installed, 0 to remove and 39 not upgraded.
Requirement already satisfied: pypandoc in /usr/local/lib/python3.7/dist-
packages (1.7.5)
Requirement already satisfied: ipyrmd in /usr/local/lib/python3.7/dist-packages
(0.4.3)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages
(from ipyrmd) (3.13)
Requirement already satisfied: nbformat in /usr/local/lib/python3.7/dist-
packages (from ipyrmd) (5.2.0)
Requirement already satisfied: traitlets>=4.1 in /usr/local/lib/python3.7/dist-
packages (from nbformat->ipyrmd) (5.1.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
/usr/local/lib/python3.7/dist-packages (from nbformat->ipyrmd) (4.3.3)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-
packages (from nbformat->ipyrmd) (4.9.2)
Requirement already satisfied: importlib-metadata in
/usr/local/lib/python3.7/dist-packages (from
jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (4.11.3)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from
jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (3.10.0.2)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-
packages (from jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (21.4.0)
Requirement already satisfied: importlib-resources>=1.4.0 in
/usr/local/lib/python3.7/dist-packages (from
jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (5.4.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/usr/local/lib/python3.7/dist-packages (from
jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (0.18.1)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/dist-
packages (from importlib-
resources>=1.4.0->jsonschema!=2.5.0,>=2.4->nbformat->ipyrmd) (3.7.0)
Collecting git+https://github.com/jupyter/nbconvert@master
  Cloning https://github.com/jupyter/nbconvert (to revision master) to /tmp/pip-
req-build-enjniftw
  Running command git clone -q https://github.com/jupyter/nbconvert /tmp/pip-
req-build-enjniftw
  WARNING: Did not find branch or tag 'master', assuming revision or

ref.
  Running command git checkout -q master
  error: pathspec 'master' did not match any file(s) known to git.
```

.3 Code informatique de traitement et d'analyse pour le calcul de la balance bénéfice risque

Le document pdf qui suit est issu de l'exécution du code Python et disponible à l'adresse https://colab.research.google.com/drive/1-m-yTOSfSZcPONa5pmmkRYEMl0wT5iYf#scrollTo=HGSncCMiev9a. Il montre les détails du code de traitement et d'analyse informatique permettant de calculer les balances bénéfices risques des vaccins Pfizer. Il est possible d'exécuter le même code en choisissant le vaccin Moderna. Nous mettons à disposition ici le code de calcul pour l'injection Pfizer (nous utilisons le même code de calcul pour l'injection de type Moderna)

Balance_Benefice_Risque_Pfizer

April 6, 2022

Etude sur la pharmaco-vigilance des vaccins Covid19

Vincent PAVAN : mathématicien, auteur, enseignant-chercheur, Aix-Marseille Université

Emmanuelle DARLES: informaticienne, enseignante-chercheuse, Université de Poitiers

1 Selection du vaccin

- **choisissez le vaccin d'étude** dans le menu déroulant juste en dessous
- cliquez sur **Execution** dans le menu en haut => **"Tout exécuter"**
- **après execution, la partie 5 présente la balance bénéfices / risques du vaccin choisi**
- la partie 4 donne des détails sur les effets indésirables graves (nombre de décès, pathologies, ...)

1.1 Date d'extraction de la base de données : 04/04/22

[88]:
```
vaccin = "Pfizer" #@param ["Pfizer", "Moderna", "Astrazeneca", "Janssen"]
recap = dict()
```

2 Introduction

Nous effectuons dans ce document des calculs statistiques liés à la pharmacovigilance des vaccins Covid19 : Pfizer, Moderna, Astrazeneca, Janssen.

Les données sont issues du site Europa mettant à disposition les cas recensés d'évènements indésirables graves et non graves dû à au moins une injection du vaccin dans l'ensemble des pays européens: https://www.adrreports.eu/fr/search_subst.html# Ces données sont réparties en plusieurs fichiers distincts: l'un pour les EI qui sont graves (Serious) et l'autre pour les fichiers qui ne sont pas graves (Unimportant).

Il n'existe pas pour la France de fichiers numériques directement téléchargeables avec lesquels on pourrait rapidement récupérer les données présentées dans l'article. Dit autrement, les personnes qui voudraient refaire des calculs et produire eux-mêmes leurs analyses des chiffres présentés sont obligé de recopier "à la main" les données, ce qui ne facilite pas le travail. On attendrait un peut plus de coopération de la part des organismes de pharmaco-vigilance.

En fait, la seule manière de rendre les choses transparentes serait de mettre à la disposition de la communauté l'ensemble des données (anonymées) permettant une analyse fiable. De ce fait, les personnes désireuses de pouvoir produire des résultats statistique le feraient avec la condition de présenter leurs calculs et leurs programmes de façon publique et reproductible. Ainsi, il apparaît que la pharmaco-vigilance est une fiche par individu. Les informations qui y sont contenues doivent être communiquée telles quelles.

Alors que les langages standardisés (PYTHON) possèdent des fonctionnalités très puissantes qui permettent la production quasi instantannée de données statistiques sur des tableaux de n'immporte quelle taille, on reste consternés par le niveau de partage et de reproductibilité qui sont fournis par la phramaco-vigilance. Pourtant l'affaire semble assez simple du point de vue informatique. Ce qu'il manque: un fichier type .csv qui pour chaque individu ayant fait remonter un effet indésirable lié à un vaccin renseigne toutes les variables possibles que les personnes .

La mise à jour de ce fichier ainsi que son partage n'ont aucune raison de ne pas exister via une mise à disposition sur un GitHub un un URL stable. On ne voit pas vraiment pourquoi un site comme OWID serait capable de mettre à disposition des données mondiales, mises à jour quotidiennement sur plus de 60 champs et pas les autorités sanitaires françaises. Il ne fait aucun doute que les bases de données de l'ANSM sont suffisamment bien construites pour qu'une exportation automatique soit produite.

3 Eléments de rappels sur l'exploitation probabiliste des résultats.

3.1 Odds ratio.

Dans cette partie, nous rappelons quelques notions utiles de probabilité. Cela nous permettra également de mieux comprendre quelles sont les formules derrière les procédures numériques que nous utilisons pour calculer les facteurs de risques déterminés par Odds Ratio.

3.1.1 Définitions et rappels

Soit p une valeur de probabilité $0 < p < 1$. On appelle Odds associée à p la valeur

$$\mathtt{O}(p) = \frac{p}{1-p} \quad \in (0, +\infty) \tag{1}$$

La traduction française du terme "Odds" se fait souvent par le mot "cote" (à entendre par exemple comme la cote associée à un pari). La cote compare la probabilité d'un évènement par rapport à son évènement complémentaire. En fait La cote est simplement une autre manière de lire une valeur de probabilité. Il y a ainsi une bijection de l'ensemble $[0, 1]$ dans l'ensemble $[0, +\infty]$ (l'inverse étant la fonction $a \mapsto a/(1 + a)$)

- ainsi une cote nulle signifie exactement une probabilité nulle
- une cote équilibrée ($\mathtt{O} = 1$) signifie une probabilité exactement de 1/2
- une cote infinie signifie un évènement certain ($p = 1$)

Cependant, cette "représentation" de la probabilité est asymétrique: tous les $p \leq 1/2$ sont dans $[0,1]$ et tous les $p > 1/2$ dans $(1,+\infty)$. Pour "symétriser" cette représentation, on définit donc le "logit" de la probabilité en prenant le logarithme du l'Odds, à savoir:

$$\texttt{logit}\left(p\right) := \ln\left(\texttt{O}\left(p\right)\right) = \ln\left(p\right) - \ln\left(1-p\right) \quad \in \left(-\infty,+\infty\right) \tag{2}$$

Ainsi les choses sont beacuoup plus symétrique: en effet si l'on pose $p = 1/2 + \pi$ alors on a

$$\forall \pi \in \left(-\frac{1}{2},\frac{1}{2}\right), \texttt{logit}\left(\frac{1}{2}-\pi\right) = -\texttt{logit}\left(\frac{1}{2}+\pi\right) \tag{3}$$

D'une manière général un "modit" est une fonction (symétrique continue strictement croissante) permettant de transformer l'intervalle $(0,1)$ en l'intervalle $(-\infty,+\infty)$

La fonction `logit` admet un inverse qui va de $(-\infty,+\infty) \mapsto (0,1)$ qui est dénommée la fonction logistique standard (également la fonciton sigmoïde), dénotée par $\sigma\texttt{logit}$ et définie par la formule:

$$\forall x \in \left(-\infty,\infty\right), \quad \sigma\texttt{logit}\left(x\right) = \frac{\exp\left(x\right)}{1+\exp\left(x\right)} \tag{4}$$

$$= \frac{1}{1+\exp\left(-x\right)} \tag{5}$$

On peut donc retrouver la valeur d'une probabilité $p \in (0,1)$ en appliquant la fonciton sigmoïde à son logit.

Une des raisons de l'introduction du logarithme de l'Odds $\texttt{O}(p)$ (le logit ou d'une manière générale ce qu'on appelle en langage informatique les "links"), c'est que cette fonction est plus facilement modélisation que la fonction $\texttt{O}(p)$ ou la fonction p elle-même.

Le `logit` nous servira effectivement dans la suite, lorsque l'on développera ce que l'on appelle des Modèles linéaires généralisés ("GLM" en anglais). Dans l'ensemble il existe énormément d'outils prêts à l'emploi dans les librairies informatiques (R, PYTHON) et donc cela permet aussi une standardisation des analyses et donc de leur comparaison.

Si p_1, p_2 sont deux valeurs de probabilités, leur Odds Ratio est défini par la relation:

$$\texttt{OR}\left(p_1,p_2\right) = \frac{\texttt{O}\left(p_1\right)}{\texttt{O}\left(p_2\right)} = \frac{p_1}{1-p_1}\frac{1-p_2}{p_2} \tag{6}$$

Le rapport des cotes ne permet de remonter directement au rapport des probabilités, mais il permet de "lire" le rapport des probabilité dans le sens suivant

- on a $\texttt{OR}\left(p_1,p_2\right) = 1 \Leftrightarrow p_1 = p_2$
- on a $\texttt{OR}\left(p_1,p_2\right) < 1 \Leftrightarrow p_1 < p_2$
- on a $\texttt{OR}\left(p_1,p_2\right) > 1 \Leftrightarrow p_1 > p_2$

En pratique, on estime plutôt le logarithme de cet OR que l'on appelle "contraste logistique".

$$\ln\left(\texttt{OR}\left(p_1, p_2\right)\right) = \texttt{logit}\left(p_1\right) - \texttt{logit}\left(p_2\right) \tag{7}$$

3.1.2 Variables aléatoires

De façon totalement mathématiques, les statistiques constituent des problèmes inverses sur les probabilités. Contrairement à une idée mal diffusée, *les probabilités nes servent pas à modéliser le hasard, mais à établir des règles de calculs rigoureux sur les mesures de sous-ensembles.* En ce sens les probabilités rejoignent le plus souvent une bonne partie de ce que l'on appelle la "statistique descriptive".

L'ensemble Ω est l'ensemble des individus que l'on veut étudier. Dans le cas de la pharmacovigilnce il s'agit de l'ensemble des personnes qui ont fait remonter un effet secondaire indésirable après une injection. Un individu de cet ensemble est noté $\omega \in \Omega$.

Les variables aléatoires sont des applications $X : \Omega \mapsto \mathcal{F}$ ***qui associe une information à un individu***. L'ensemble d'arrivé $\mathcal{F}$ peut être un ensemble numérique (i.e. un sous ensemble des réels) ou pas.

- Si l'ensemble d'arrivée est inclus dans $\mathbb{R}$ on fit que la variable aléatoire est quantitative
- Si l'ensemble d'arrivée n'est pas un ensemble de nombre réels, on dit que la variable alatoire est qualitative. Le "codage" est alors ce qui permet de transformer une variable aléatoire qualitative en
- Même lorsqu'il n'est pas un ensemble de nombre, l'ensemble $\mathcal{F}$ peut être un ensemble ordonné (par exemple des classes d'âges), c'est à dire un ensemble possédant un ordre (total). Dans ce cas la variable aléatoire est dite ordinale. Les variables aléatoires ordinales permettent le calcul d'événènement dits "cumulés".

Le cas le plus simple est celui où la variable aléatoire ne peut prendre que deux valeurs. On parle alors de variable aléatoire binomiale.

Pour les données de pharmacovigilance qui sont données, une variable aléatoire que l'on peut associer à chaque individu est celle de "la gravité". Ainsi on considérera la variable Y définie par:

$$Y : \Omega \mapsto \{0, 1\} \tag{8}$$
$$\omega \mapsto Y\left(\omega\right) \tag{9}$$

Avec $Y\left(\omega\right) = 1$ si l'EI (effet indésirable) est grave et 0 si l'EI n'est pas grave. En pratique l'idée c'est de chercher à savoir si cette variable peut ou non "être exliquée" par d'autres variables. Ces autres variables on les appelles les "variables explicatives". En pratique pour nous ce seront les variables de type sexe, âges, lieu de résidence (EHPAD), comorbidité (hypertension, diabète) etc. On notera par $X_i : i \in [1, K]$ ces variables. Elles peuvent être de type:

- binomiales: par exemple le sexe
- continue (comme par exemple l'âge)
- ordinale: les individus sont ramenés à une classe d'âge

- descriptive: la nom de la pathologie associée à l'effet indésirable dans la classification SOC (System Organ Class)
- etc.

Pour un sous-ensemble $\mathcal{A} \subset \mathcal{F}$ où $\mathcal{F}$ est l'ensemble image de Ω par la variable aléatoire X, on appelle évènement associé à $\mathcal{A}$ le sous-ensemble de Ω obtenu par *l'image inverse* de la variable aléatoire X:

$$\forall \mathcal{A} \subset \mathcal{F}, \quad X^{-1}(\mathcal{A}) = \{\omega \in \Omega, \quad X(\omega) \in \mathcal{A}\} \tag{10}$$

On note par $\{X = \mathcal{A}\}$ ou simplement $X = \mathcal{A}$ un tel sous ensemble.

Dans la suite, les probabilités d'intérêts seront décrites par des probabilités conditionnelles pour des évènements attachés à des variable aléatoire. Ainsi, ce que l'on cherchera à modéliser sera le plus souvent

$$p := P(Y = y \mid X = x) \tag{11}$$

Ainsi, si Y est la variable aléatoire codant le fait d'être mort ou non du vaccin, et si X code le sexe, alors $P(Y = 1 \mid X = x)$ détermine la probabilité de mort sachant que l'on a le sexe x. Le rapport des côtes permet ainsi de savoir si ces

3.1.3 Probabilité d'intérêt

Pami toutes les variables aléatoires attachées à la pharmaco-vigilance d'un produit, on va chercher à "exliquer" certaines variables (par exemple la gravité de l'EI) par d'autres variables (par exemple le sexe et l'âge).

- les variables à expliquer sont les variables "dépendantes"
- les variables explicatives sont les "variables indépendantes"

Par exemple on voudra savoir si le fait d'être une femme "explique" la gravité ou non de l'EI au sens suivant: une femme a-t-elle plus de "chances" qu'un homme d'avoir un effet indésirable grave (lorsqu'elle a un EI) ? Ainsi posée, la question n'a pas de formulation mathématiques. POur y arriver il faut affiner la question: chez les femmes, le rapport entre celles ayant un EI grave et celles ayant un EI non grave est-il supérieur au rapport entre les hommes ayant eu un effet grave et ceux ayant eu un effet indésirable non grave ?

On note par $Y \in \{G, NG\}$ la variable aléatoire qui quantifie la gravité (pour nous c'est la variable Z_7 du DataFrame précédent). Elle est de type binomiale, c'est à dire qu'elle ne peut prendre que deux modalités. On note par $X \in \{H, F\}$ la variable qui code le sexe de la personnes (pour nous c'était la variable Z_1 du DataFrame précédent). On va reconstruire pour chacune des modalités des variables

3.2 Modèles linéaires généralisées et calcul d'OR

3.2.1 Le principe des GLM

Etant donnée une variable d'intérêt Y (par exemple un variable binomiale qui renseigne sur la gravité ou pas de l'EI secondaire, le fait qu'il conduise ou non à la mort etc.), on a vu que l'on s'intéressait fondamentalement à la détermination de la quantité:

$$\forall \left(x_1, \cdots, x_K\right), \; P\left(Y = y_i \mid X_1 = x_1 \wedge \cdots \wedge X_K = x_K\right) \tag{12}$$

comme on a fixé la valeur y_i de l'évènement Y, on voit que la probabilité qui nous intéresse est uniquement une fonction des valeurs $(x_1, \cdots, x_K)$ prises *simultanément* par les variables aléatoires $(X_1, \cdots, X_K)$. La modélisation consiste à trouver une fonction permettant le calcul suivant

$$P\left(Y = y_i \mid X_1 = x_1 \wedge \cdots \wedge X_K = x_K\right) := f_i\left(x_1, \cdots, x_K\right) \tag{13}$$

L'intérêt d'un tel "modèle", c'est que si on se donne ensuite une valeur particulière du K-uplet $(x_1, \cdots, x_K)$ on peut alors calculer rapidement la probabilité d'intérêt: cela donne, pour un individu qui vérifie les conditions:

$$X_1\left(\omega\right) = x_1 \wedge \cdots \wedge X_K\left(\omega\right) = x_K \tag{14}$$

la probabilité qu'il expérimente l'évènement $\{Y = y_i\}$. C'est ce qu'on appelle la prédiction.

Il existe des méthodes assez systématiques pour pouvoir trouver la fonction f_i.

1. D'abord on cherche quelle est la forme de la variable Y à expliquer. Selon sa forme (binomiale, normale, Poisson etc.) et selon ce que l'on connaît des variables aléatoires $(X_1, \cdots, X_K)$, on va appliquer à la variable:

$$P\left(Y = y_i \mid X_1 = x_1 \wedge \cdots \wedge X_K = x_K\right) \tag{15}$$

une transformation de type "modit":

- la fonction `logit`
- la fonction `probit`
- etc.

Dans le langage informatique, la transformation "modit" est dénommée la fonction "link".

2. Une fois que l'on a choisi la fonction" l'hypothèse GLM (generalized Linearized Model) consiste à trouver une fonction linéaire pour "fittter" la transformée "modit" de la probabilité. Concrètement on cherche des coefficients linéaires $(\beta_0, \cdots, \beta_K)$ tels que l'on ait:

$$\texttt{modit}\left(P\left(Y = y_i \mid X_1 = x_1 \wedge \cdots \wedge X_K = x_K\right)\right) = \beta_0 + \sum_{l=1}^{l=K} \beta_l x_l \tag{16}$$

La manière de trouver les coefficients linéaires $(\beta_l : l \in [0, K])$ dépent du type de la variable Y. C'est l'estimateur du Maximum de vraisemblance qui est retenu pour la détermination des coefficients linéaires et ce dernier dépend de la loi suivie par Y.

3.2.2 Codage des variables qualitatives

Un des problèmes de la présentation précédente, c'est que les variables d'intérêt ne sont en général pas des variables de type quantitatif:

1. Ainsi, si l'on imagine que pour la variable Y on ne considère qu'une seule variable explicative X_1 (par exemple le sexe) celle-ci ne prendra que les valeurs $X_1 \in \{F, H\}$ qui ne sont éviemment pas numériques.
2. Si l'on imagine que la variable Y doit être expliquée par la classe d'âge X_2 qui peut désigner L intervalles différents (et les intervalles ne sont pas en tant que tels des nmbres) qui seront notés symboliquement $X_2 \in \{I_1, \cdots, I_L\}$

D'une manière générale, on désignera par "modalité" une valeur possible prise par une variable $X_k : k \in [1, K]$. Lorsque ces modalités sont en nombre finis J_k, on va créer en fait $J_k - 1$ variables binomiales ($j \in [1, J_k - 1]$) de la façon suivante:

$$\forall k \in [1, K], \quad \forall j \in [1, J_k - 1], \quad X_k(\omega) = x_k \Leftrightarrow X_k^j(\omega) = 1 \tag{17}$$
$$\forall k \in [1, K], \quad \forall j \in [1, J_k - 1], \quad X_k(\omega) \neq x_k \Leftrightarrow X_k^j(\omega) = 0 \tag{18}$$

Exemple: Supposons que la variable X_2 soit la variable de la classe d'âge. On suppose que ce classes d'âges sont les modalités qui suivantes:

$$I_0 = [12, 17], \ I_1 = [18, 64], I_2 = [65, 85], \ I_3 = [85+) \tag{19}$$

On a donc $J_2 = 4$ et on va créer $J_2 - 1 = 3$ variables binomiales qui sont X_2^1, X_2^2, X_2^3 définies respectivement par:

$$X_2^1(\omega) = 1 \Leftrightarrow X_2(\omega) = I_1, \quad X_2^1(\omega) = 0 \Leftrightarrow X_2(\omega) \neq I_1 \tag{20}$$
$$X_2^2(\omega) = 1 \Leftrightarrow X_2(\omega) = I_2, \quad X_2^2(\omega) = 0 \Leftrightarrow X_2(\omega) \neq I_2 \tag{21}$$
$$X_2^3(\omega) = 1 \Leftrightarrow X_2(\omega) = I_3, \quad X_2^3(\omega) = 0 \Leftrightarrow X_2(\omega) \neq I_3 \tag{22}$$

Ainsi, si l'on cherche à expliquer la variable $Y = y_i$ en fonction de la variable X_k qui prend J_k modalités, on écrira que l'on a:

$$\texttt{modit}\left(P\left(Y = y_i \mid X_k\right)\right) = \beta_0 + \sum_{j=1}^{J_k - 1} \beta_j X_k^j \tag{23}$$

La modalité 0 s'appelle la modalité de référence. En pratique c'est un choix arbitraire. Il dépend de ce que souhaite l'utilisateur dans la descritpion des probabilités. On pourra donc revenir à la probabilité $P(Y = y_i \mid X_k)$ par application de la fonction réciproque $\sigma\texttt{modit}(\cdot)$ de la manière suivante:

$$P\left(Y = y_i \mid X_k\right) = \sigma\texttt{modit}\left(\beta_0 + \sum_{j=1}^{J_k - 1} \beta_j X_k^j\right) \tag{24}$$

Ainsi, on a deux cas de figure

1. Si l'on vérifie $X_k(\omega) = I_0$, on sait alors que l'on a $X_k^j(\omega) = 0$ pour tous les $j \in [1, J_K - 1]$. On en déuit alors:

$$P(Y = y_i \mid X_k = I_0) = \sigma\mathtt{modit}(\beta_0) \tag{25}$$

2. Si l'on vérifie $X_k(\omega) = I_j$, on sait alors que l'on a $X_k^j(\omega) = 1$ et que tous les autres $X_k^l(\omega)$ sont nuls pour $l \neq j$. On en déduit alors:

$$P(Y = y_i \mid X_k = I_l) = \sigma\mathtt{modit}(\beta_0 + \beta_l) \tag{26}$$

3.2.3 Recouvrement des OR

Une fois connu les paramètres du GLM, on possède donc une méthode simple pour retrouver le calcul des OR. Pour cela il suffit de voir que l'on a la formule

$$\ln((\mathtt{OR}(p_1, p_2))) = \mathtt{logit}(p_1) - \mathtt{logit}(p_2) \tag{27}$$

Par conséquent, pour trouver la valeur des `OR` il suffit de prendre l'exponentiel de la différence des `logit`.

Imaginons par exemple que, pour le risque Y qui est celui de la gravité, a obtenu les coefficients du GLM pour la variable explicative (le risque) qui est celle du sexe X_1. On suppose que la variable X_1 (qui peut prendre uniquement deux modalités) est codée de sorte que $X_1^1(\omega) = 0$ si $X_1(\omega) = F$ et $X_1^1(\omega) = 1$ si $X_1(\omega) = M$. On écrit ainsi

$$\mathtt{logit}(P(Y = S \mid X_1)) = \beta_0 + \beta_1 X_1^1 \tag{28}$$

On peut alors écrire:

$$\begin{aligned}
\ln(\mathtt{OR}(P(Y = S \mid X_1 = F), P(Y = S \mid X_1 = M))) &= (\beta_0) - (\beta_0 + \beta_1) && (29)\\
&= -\beta_1 && (30)\\
\mathtt{OR}(P(Y = S \mid X_1 = F), P(Y = S \mid X_1 = M)) &= \exp(-\beta_1) && (31)
\end{aligned}$$

De la même manière, on peut calculer les `OR` lorsque l'on connaît les coefficients β d'un GLM.

On rappelle que sur l'ensemble Ω, la probabilité choisie est toujours la probabilité uniforme: pour avoir la probabilité d'un sous-ensemble $A \subset \Omega$ on effectue le rapport:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} \tag{32}$$

où $\mu(\cdot)$ est la fonction de cardinal, i.e. à un sous-ensemble elle associe son nombre d'éléments.

Si B est un évènement (i.e. un sous ensemble) de Ω, de probabilité non nulle, on appelle probabilité conditionnelle (sachant B) la probabilité suivante sur l'ensemble Ω

$$P_B(A) := \frac{P(A \bigcap B)}{P(B)} := P(A \mid B) \tag{33}$$

Les événements peuvent-êtres décrits par des variables aléatoires. En ce qui nous concerne, l'évènement B sera celui qui sera écrit par

$$B = \{X_1 = x_1, X_2 = x_2, \cdots\} \tag{34}$$

où X_i est une variables aléatoires "explicative" (classe d'âge, sexe, etc.) et x_i une valeur (ou de façon équivalente dans le vocabulaire une modalité) de l'ensemble $\mathcal{F}_i$ (avoir tel âge, être un homme etc.).

Dans les modèles dits "logit", ce qui nous intéresse, c'est la probabilité suivante:

$$P(Y = 1 \mid (X_1 = x_1, X_2 = x_2, \cdots)) \tag{35}$$

Un modèle linéaire généralisé (glm en anglais: generalized linear model) c'est ce qui permet d'estimer le logit de la garendeur $P(Y = 1 \mid (X_1, X_2, \cdots))$ sous forme d'une fonction linéaire (ou affine) de $X_1, X_2, \cdots$ (qui est alors nécessairement codée sous forme numérique). On pose ainsi:

$$\texttt{logit}(P(Y = 1 \mid (X_1, X_2, \cdots))) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \tag{36}$$

En inversant la fonction $\ln(x/(1-x))$ qui est la fonction $1/(1+\exp(-x))$ (que l'on appelle la fonction sigmoïde logisitique) que l'on note par $\texttt{logit}^{-1}$ (et que l'on notera par $\sigma\texttt{logit}$ on a

$$P(Y = 1 \mid (X_1, X_2, \cdots)) = \sigma\texttt{logit}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots) \tag{37}$$

Dans le vocabulaire de l'épidémiologie les variables Y désignent les risques tandis que les variables $(X_k : k \in [1, K])$ désignent les facteurs

4 Extraction des données

4.0.1 Notion de dataframe

Le DataFrame est une classe objet (au sens informatique du terme) existant dans de nombreux langages actuels (Python, R, etc.) et pour lequel on possède des constructeurs et des méthodes de haut niveau permettant un manipulation fiable des données qui le constituent.

Le DataFrame d'une pharmaco-vigilange est ainsi un tableau à double entrée:

- chaque ligne correspond à un individu ω. Ces individus sont repérés par un indice entier $i \in [1, N]$ où N est le nombre total des individus présents dans le DataFrame de la Pharmaco-vigilance, c'est à dire le nombre total des individus ayant été affecté par un effet indésirable.
- chaque colonne représente une variable aléatoire associé à un individu.
- On remplit le DataFrame en donnant pour chaque individu la modalité de la variable aléatoire.

La pharmaco-vigilance consiste donc au remplissage, pour chaque individu des champs contenant les informations relatives à sa propre personne et aux effest indésirables qu'il a endurés. Toutes ces informations sont exactement les variables aléatoires de la théorie. *Il y a donc autant de variable aléatoire qu'il y a de colonnes dans le DataFrame.* On peut aller voir le site du gouvernement pour avoir une description synthétique du mode de fonctionnement de la pharmaco-vigilance. Cela consiste à remplir les tableaux au fur et à mesure. Une pharmaco-vigilance est relié à un produit particulier (par exemple en ce qui nous intéresse, le vaccin Pfizer)

1. ***S'agissant de la pharmaco-vigilance européeenne, un tel DataFrame peut être obtenu, de façon séparé pour les effets graves et les effets non grave.***
2. ***Pour la France, il n'y a pas de possibilité d'aller chercher le DataFrame qui est utilisé par les agences. Les rapports qui sont fournis. Pire il n'y a pas - à notre connaissance - de moyen d'aller chercher des tableaux de données: il faut "recopier à la main" les données fournies par les rapports***

4.0.2 Localisation et extraction des données

Les données sont issues du site Europa mettant à disposition les cas recensés d'évènements indésirables graves et non graves dû à au moins une injection du vaccin d'étude dans l'ensemble des pays européens: https://www.adrreports.eu/fr/search_subst.html#

Ces données sont réparties en plusieurs fichiers distincts: l'un pour les EU qui sont graves (Serious) et l'autre pour les fichiers qui ne sont pas graves (Non Serious). Le chunk suivant (i.e. l'unité de code suivante) permet d'aller télécharger les fichiers. Ces fichiers sont ensuite fusionnés en ajoutant une nouvelle variable aléatoire qui est celle de l'indice de gravité. Ainsi, on pourra tester des Odds ratio liés à la gravité.

[141]:
```
!pip install -U -q PyDrive

import pandas
import matplotlib.pyplot as plt
from google.colab import drive
from google.colab import files
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

if vaccin == "Pfizer" :
  id_file =␣
 ↪["1FrphRsc-xjhsQ04G6Jo1_dWPgXzb5A2O","1Dqk2386vQUXYe0r1irhpeMTZX0r1Kljn"]
if vaccin == "Moderna":
  id_file =␣
 ↪["1YswqSfx1WTjXGmAhRLnxt9UOaS2broO5","1-fJsc8GL-6RruUzx9LuxStgDXn2iCpOp"]
if vaccin == "Astrazeneca":
  id_file =␣
 ↪["1vQSdZSJJNZyUt4YgCgvNGElUbYlCDoL5","1o3rZZWuNpFQOYxovKDyGYbbx072q9uXv"]
if vaccin == "Janssen":
```

```
  id_file =␣
↪["1hpc9ANVQzLTF7wv-u9Op68b0mCyYOm6N","1Yi2sfW36hhiivikCs45LPuzypeqV03mZ"]


# Montage du drive et création des répertoires
drive.mount('/content/drive')
!mkdir /content/drive/MyDrive/MetaAnalyseVaccin/
!mkdir /content/drive/MyDrive/MetaAnalyseVaccin/datas/
nameFile = ["serious.xlsx","nonSerious.xlsx"]

auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

# Téléchargement des fichiers de données et versement sur le drive
path_file_data = "/content/drive/MyDrive/MetaAnalyseVaccin/datas/"
for i in range(0,len(id_file)):
  downloaded = drive.CreateFile({"id":id_file[i]})
  downloaded.GetContentFile(path_file_data+nameFile[i])

df0 = pandas.read_excel(path_file_data+nameFile[0])
df1 = pandas.read_excel(path_file_data+nameFile[1])
output.clear()
```

[142]:
```
# On rajoute la variable aléatoire de la gravité pour chaque DataFrame et on␣
↪les fusionne
new_df0=df0.assign(Severity = "Serious")
new_df1=df1.assign(Severity = "Unimportant")
df2 = pandas.concat([new_df0, new_df1])
```

4.0.3 Variables aléatoires

Grâce aux fonctionalités du langage Python, on peut aller chercher très efficacement de nombreuses informations sur les données contenues dans le DataFrame. Cela illustre parfaitement l'intérêt d'aller travailler avec les méthodes que nous utilisons dans cette étude

[143]:
```
listeVA = pandas.DataFrame(df2.columns)
listeVA.columns = ["Information Variable aléatoire"]
ind = []
for i in range(0, len(listeVA)):
  ind.append("Z_"+ str(i))
listeVA.index = ind
listeVA.T
```

[143]:

```
                                             Z_0              Z_1          Z_2  \
Information Variable aléatoire  Unnamed: 0  EU Local Number  Report Type

                                                        Z_3  \
Information Variable aléatoire  EV Gateway Receipt Date

                                                             Z_4  \
Information Variable aléatoire  Primary Source Qualification

                                                                               Z_5
\
Information Variable aléatoire  Primary Source Country for Regulatory Purposes

                                                     Z_6                Z_7  \
Information Variable aléatoire  Literature Reference  Patient Age Group

                                                                    Z_8  \
Information Variable aléatoire  Patient Age Group (as per reporter)

                                                    Z_9         Z_10  \
Information Variable aléatoire  Parent Child Report  Patient Sex

Z_11  \
Information Variable aléatoire  Reaction List PT (Duration - Outcome -
Serious…

Z_12  \
Information Variable aléatoire  Suspect/interacting Drug List (Drug Char -
Ind…

Z_13  \
Information Variable aléatoire  Concomitant/Not Administered Drug List (Drug
C…

                                      Z_14      Z_15
Information Variable aléatoire  ICSR Form  Severity
```

La lecture des variables aléatoires doit êre documentée. Dans le cadre de la pharmacovigilance européenne, il s'agit de comprendre le fonctionnement de la plateforme **Eudravigilance**. Une thèse sur le sujet peut être trouvée à l'édresse suivante:

https://dumas.ccsd.cnrs.fr/dumas-01957354/document

L'individuation des données est assurées par deux types de variables (i.e. ces variables prennent nécessairement des valeurs distinctes pour chaque individu)

1. La première est la variable ICSR (individual Case Safety Report). Cette variable renvoie sur un lien URL dans lequel on peut retrouver les informations relatifs au remplissage de la pharmaco-vigilance.

2. La seconde est la variable EU Local Number

Pour chaque variable aléatoire, on est capable de donner l'ensemble de ses valeurs, c'est à dire l'ensemble de ses modalités. Ainsi pour la variable aléatoire de l'age on peut voir qu'il y a les modalités suivantes:

[144]:
```
pandas.DataFrame( df2.groupby(["Patient Age Group"]).indices.keys(), columns =␣
↪["Patient Age Group"])
```

[144]:
```
     Patient Age Group
0           0-1 Month
1         12-17 Years
2         18-64 Years
3  2 Months - 2 Years
4          3-11 Years
5         65-85 Years
6  More than 85 Years
```

Par défaut Python classe les modalités des variables aléatoires par ordre alphabétique. On va renommer les classes d'âges de sorte que l'ordre lexicographique corresponde à l'ordre chronologique

[145]:
```
str1 = d={"0-1 Month": "Month 0-1", "2 Months - 2 Years": "Month 2 - Years 2"}
df2 = df2.replace(d)
d = {"3-11 Years" : "Years 03-11", "12-17 Years": "Years 12-17", "18-64 Years" :
↪ "Years 18-64"}
df2 = df2.replace(d)
d = {"65-85 Years" : "Years 65-85", "More than 85 Years" : "Years 85 and More",␣
↪"Not Specified": "Yet Not Specified"}
df2 = df2.replace(d)
```

On peut également rapidement obtenir les effectifs associés à chacune des modalités des variables aléatoires. Pour cela il suffit d'avoir une variable aléatoire qui renvoie une valeur spécifique à un individu. En ce qui nous concerne, il y a deux variables aléatoires spécifiques: la variable "ICSR Form" et la vairable "EU Local Number".

[146]:
```
df2.groupby(["Patient Age Group"]).nunique()["ICSR Form"]
```

[146]:
```
Patient Age Group
Month 0-1               427
Month 2 - Years 2      1361
Years 03-11            3591
Years 12-17           23844
Years 18-64          586657
Years 65-85          156180
Years 85 and More     25230
Name: ICSR Form, dtype: int64
```

On peut évidemment trouver le total N des cas rapportés en faisant la somme des effectifs:

[147]:
```
print("Nombre total d'individus ayant eu un EI:", df2.groupby(["Patient Age␣
↪Group"]).nunique()["ICSR Form"].sum())
print("Nombre total d'individus ayant eu un EI grave:", new_df0.
↪groupby(["Patient Age Group"]).nunique()["ICSR Form"].sum())
```

```
Nombre total d'individus ayant eu un EI: 797290
Nombre total d'individus ayant eu un EI grave: 249103
```

On peut également retrouver les effectifs en effectuant l'intersection des variables aléatoires. Dans le formalisme que l'on a choisi, cela revient à rechercher les

$$\texttt{Card}\left(Y_1 = y_1, X_1 = x_1, X_2 = x_2\right) \tag{38}$$

où les (y_1, x_1, x_2) prennent toutes les valeurs possibles dans $\{NS, S\} \times \{F, M\} \times \{I_0, \cdots, I_7\}$

[148]:
```
dfHP = df2[df2["Primary Source Qualification"]=="Healthcare Professional"]
dfHP.groupby(["Patient Sex","Severity"]).nunique()["ICSR Form"]
```

[148]:
```
Patient Sex        Severity
Female             Serious        75726
                   Unimportant    99017
Male               Serious        46348
                   Unimportant    39872
Yet Not Specified  Serious         1153
                   Unimportant     2794
Name: ICSR Form, dtype: int64
```

Il va de soit que l'on conserve exactement le nombre total de cas: cela signifie que l'on a bien effectué une partition de l'ensemble.

[149]:
```
print(df2.groupby(["Patient Age Group"]).nunique()["ICSR Form"].sum() == df2.
↪groupby(["Patient Age Group", "Patient Sex","Severity"]).nunique()["ICSR␣
↪Form"].sum())
```

```
True
```

4.1 Fiabilité des données

Pour chaque personne, il est également possible de savoir si la remontée des EI a été réalisée par un personnel soignant ou un non soignant. La colonne "Primary Source Qualification" possède 2 valeurs différentes : - "Healthcare Professionnal" : la déclaration des EI pour la personne concernée a été faite par un soignant; - "Non Healthcare Professional" : cette déclaration a été faite par un non soignant.

Il est donc possible de voir en quelle proportion les remontées ont été réalisées.

[150]:
```
dfHP = df2[df2["Primary Source Qualification"]=="Healthcare Professional"]
dfPers = df2[df2["Primary Source Qualification"]=="Non Healthcare Professional"]
```

```
print("Nombre de cas déclarés par un personnel soignant : "+str(len(dfHP)))
print("Nombre de cas déclarés par un non soignant : "+str(len(dfPers)))
```

```
Nombre de cas déclarés par un personnel soignant : 274282
Nombre de cas déclarés par un non soignant : 538013
```

4.2 Exploitation numérique des résultats

Maintenant que nous avons reproduit les données, nous pouvons masser méthodiquement à la production de calculs d'après les chiffres qui ont été fournis.

4.3 Principe de comparaison

Dans l'avaluation d'un vaccin qui doit protéger d'une maladie, il s'agit essentiellement de mettre en place la construction systématique d'une balance bénéfice/risque. Cette balance doit être construite de façon méthodique en fonction de chaque type de maladie et de chaque type de vaccin. Le principe général que nous posons ici est le suivant

vaccin* ∼ *maladie

Il faut ensuite donner une façon de transposer tous les indicateurs liés à la maladie sur les indicaturs liés au vaccin.

1. ***Première Transposition: Effet Indésirable* ∼ *Maladie Symptomatique***
2. ***Deuxème Transposition: Mort post vaccinale* ∼ *Mort de la Maladie***
3. ***Troisième transposition: Hospitalisation pour vaccin* ∼ *Hospitalisation pour Covid***
4. ***Quatirème transposition: Létalité apparente pour vaccin* ∼ *Létalité apparente pour Covid***
5. etc.

Une difficulté importante, c'est qu'il existe sur les vaccins un indicateur de gravité qui n'existe pas en général sur la maladie.

La transposition ainsi construite doit pouvoir être abordée de façon absolue et de façon relative

1. De façon absolue: il s'agit de donner par exemple de définir des taux de mortalités et des taux de létalités pour les vaccins et pour les maladies, à la fois en populaiton générale, par classe d'âge etc... Un des gros problèmes liés à cette comparaison est la suivante: on sait que les EI ne font l'bjet que d'une remontée imcomplète. Ainsi, sur tous les indicateurs absolus liés à la pharmaco-vigilance, il faudra appliquer a priori un facteur multiplicatif représentant les effets de sous-déclaration. A l'inverse, on sait que les indicateurs absolus liés au COVID (en particulier les morts) peuvent être sur-déclarés (mort avec le COVID vs mort à cause du COVID). Il faudra donc toujours tenir compte de ces décalages dans les comparaision.
2. *De façon relative: il s'agit de la comparaison des catégories entre elles, autrement dit d'établir les facteurs de risque sur des critères donnés.* Par exemple parmi les gens qui meurent (du vaccin ou de la maladie) on compare les facteurs de risque sur les classes d'âges ou sur les

sexes. Pour ces comparaisons relatives et pour tenir compte du fait qu'il n'existe pas, pour le COVID, d'indicateur sur l'échelle de gravité.

La locution "facteurs de risque" doit être décomposée selon deux principes: celui du "risque" et celui du "facteur". Dans le formalisme que nous avons posé, le "risque" est celui qui est d'expérimenter une certaine valeur d'une variable aléatoire de type Y: il s'agit de la variable (dépendante) à expliquer. Dans le même formalisme qui a été présenté les "facteurs" constituent l'ensemble des valeurs prises par les variables aléatoires de type $(X_k : k \in [1, K])$

1. Pour la vaccination, il y a deux deux risques que nous allons définir

- le risque de l'effet grave
- le risque de décès

Du point de vue de la formalisation, pour chaque individu subissant un EI, cela revient à définir les deux variables aléatoires Y_1, Y_2 suivantes:

$$\forall \omega \in \Omega, \quad Y_1(\omega) = \texttt{Serious} \Leftrightarrow \textit{l'EI est grave} \qquad (39)$$
$$\forall \omega \in \Omega, \quad Y_1(\omega) = \texttt{Unimportant} \Leftrightarrow \textit{l'EI est non grave} \qquad (40)$$

Ainsi que la variable liée aux décès:

$$\forall \omega \in \Omega, \quad Y_2(\omega) = \texttt{Fatal} \Leftrightarrow \textit{la personne décède} \qquad (41)$$
$$\forall \omega \in \Omega, \quad Y_2(\omega) = \texttt{Survive} \Leftrightarrow \textit{la personne survit} \qquad (42)$$

2. Parmi les facteurs, nous nous concentrerons essentiellement sur deux facteurs: la variable X_1 qui donne le sexe et la variable X_2 qui donne l'âge. La première est de type binomiale, tandis que la seconde est de type polynomiale avec des classes d'âge. On a donc:

$$\forall \omega \in \Omega, \quad X_1(\omega) = \texttt{F} \Leftrightarrow \textit{la personne est une femme} \qquad (43)$$
$$\forall \omega \in \Omega, \quad X_1(\omega) = \texttt{M} \Leftrightarrow \textit{la personne est un homme} \qquad (44)$$

Pour la variable de classe d'âge nous aurons:

$$\forall \omega \in \Omega, \quad X_2(\omega) = I_j \Leftrightarrow \textit{la personne est dans la calsse d'âge } j \qquad (45)$$

Le but c'est de comparer esuite les OR sur les facteurs de risque à la fois pour la maladie et pour la vaccination.

Une manière possible d'évaluer l'adéquation du vaccin à la maladie consitera finalement à définir un score comme étant par exemple un coefficient de corrélation global entre les indicateurs liés au vaccin et ceux liés à la maladie.

4.4 Exploitation numérique à partir du dataframe européen : indicateurs généraux

Comptage des effets "Graves" / "Bénins" en fonction du sexe et du groupe d'âge

[151]:
```
from IPython.display import display, HTML
%load_ext google.colab.data_table
from google.colab import data_table
features = {"Patient Sex" : ["Female","Male"],
            "Patient Age Group":["0-1 Month","2 Months - 2 Years", "3-11␣
↪Years","12-17 Years", "18-64 Years", "65-85 Years","More than 85 Years"]}
dfHP0 = df0#[df0["Primary Source Qualification"]=="Healthcare Professional"]
dfHP1 = df1#[df1["Primary Source Qualification"]=="Healthcare Professional"]

dfFeature = dict()
for key in features.keys():
  values = features[key]
  for value in values :
    c1 = len(dfHP0[dfHP0[key] == value])
    c2 = len(dfHP1[dfHP1[key] == value])
    if c2!=0 :
      ratio = round((c1/c2),2)
    else : ratio = None
    dfFeature.update({value:[c1,c2,ratio]})

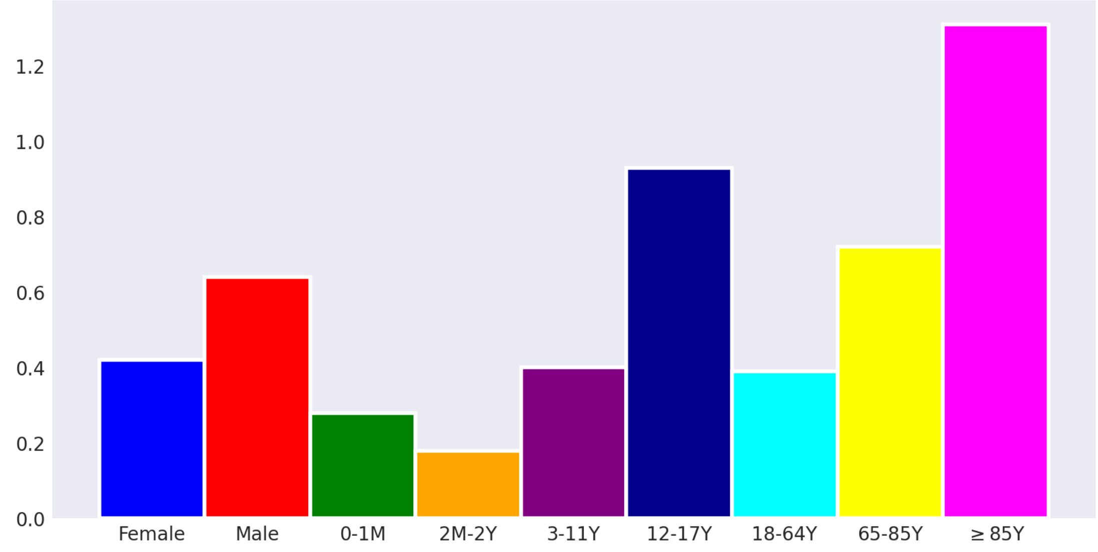
dfFeature = pandas.DataFrame.from_dict(dfFeature)
dfFeature = dfFeature.transpose()
dfFeature.rename(columns={0: 'Serious', 1:"Unimportant", 2:
↪"ratio"},inplace=True)
dfFeature.to_excel(path_file_data+'table.xlsx',sheet_name='sheet1',index=True)
y = dfFeature["ratio"]
data_table.DataTable(dfFeature)
strg_EI_euro = dfFeature.to_markdown()
print(strg_EI_euro)
```

```
The google.colab.data_table extension is already loaded. To reload it, use:
  %reload_ext google.colab.data_table
|                    |   Serious |   Unimportant |   ratio |
|:-------------------|----------:|--------------:|--------:|
| Female             |    169410 |        400704 |    0.42 |
| Male               |     91104 |        142268 |    0.64 |
| 0-1 Month          |        93 |           334 |    0.28 |
| 2 Months - 2 Years |       207 |          1154 |    0.18 |
| 3-11 Years         |      1027 |          2564 |    0.4  |
| 12-17 Years        |     11466 |         12378 |    0.93 |
| 18-64 Years        |    165158 |        421499 |    0.39 |
| 65-85 Years        |     70699 |         98492 |    0.72 |
| More than 85 Years |     15458 |         11766 |    1.31 |
```

[152]:
```
fig = plt.figure()
fig.set_size_inches(20, 10)
ax = plt.subplot(1,1,1)
```

```
width = 0.2
x = [i for i in range(1,10)]
ax.bar(1,dfFeature["ratio"][0],width = 1, color="blue",linewidth = 4,␣
↪align='center',label="Female")
ax.bar(2,dfFeature["ratio"][1],width = 1, color="red",linewidth = 4,␣
↪align='center',label="Male")
ax.bar(3,dfFeature["ratio"][2],width = 1, color="green",linewidth = 4,␣
↪align='center',label="0-1 Month")
ax.bar(4,dfFeature["ratio"][3],width = 1, color="orange",linewidth = 4,␣
↪align='center',label="2 Months - 2 Years")
ax.bar(5,dfFeature["ratio"][4],width = 1, color="purple",linewidth = 4,␣
↪align='center',label="3-11 Years")
ax.bar(6,dfFeature["ratio"][5],width = 1, color="darkblue",linewidth = 4,␣
↪align='center',label="12-17 Years")
ax.bar(7,dfFeature["ratio"][6],width = 1, color="cyan",linewidth = 4,␣
↪align='center',label="18-64 Years")
ax.bar(8,dfFeature["ratio"][7],width = 1, color="yellow",linewidth = 4,␣
↪align='center',label="65-85 Years")
ax.bar(9,dfFeature["ratio"][8],width = 1, color="magenta",linewidth = 4,␣
↪align='center',label="More than 85 Years")
plt.xticks(x,fontsize="20")
ax.
↪set_xticklabels(["Female","Male","0-1M","2M-2Y","3-11Y","12-17Y","18-64Y","65-85Y","$\geq85
plt.grid()
plt.yticks(fontsize="20")
plt.title("Ratio entre le nombre d'effets indésirables graves et non␣
↪graves",fontsize="25",weight='bold')
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/ratio_"+vaccin+".png")
# plt.savefig("ratio.png")
```





Liste des pathologies ayant entraînées le plus d'hospitalisation Afin d'extraire les pathologies qui ont entraînées des hospitalisation et qui émergent des données, nous calculons les fréquences d'apparition de ces pathologies

[153]:
```
import numpy as np
from IPython.display import display, HTML

def parseReaction(reaction):
  reaction = list(reaction)
  for i in range(0,len(reaction)):
    if reaction[i]=="(":
      j = i
      while reaction[j]!=")":
        reaction[j]=""
        j+=1
      reaction[j]=""
    if reaction[i]=="<":
      j = i
      while reaction[j]!=">":
        reaction[j]=""
        j+=1
      reaction[j]=""
  reaction = "".join(reaction)
  reaction = reaction.split(",")
  return reaction

def computeFreq(search):
  df = dict()
  reactions = df0['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)']
  for index, value in reactions.items():
    reaction = str(value)
    reaction = reaction.split("<BR><BR>")
    for r in reaction :
      if r.find(search)!=-1:
        r = parseReaction(r)
        for re in r :
          if re!="" and re!=" ":
            if search=="Fatal":
              if re.find("Death")!=-1 or re.find("Sudden death")!=-1:
                re = "Unknown (death or sudden death)"
            if re in df.keys():
              df[re][0]+=1
            else :
```

```
            df.update({re:[1]})
  dfPatho = pandas.DataFrame.from_dict(df)
  nbPatho = len(dfPatho.keys())
  dfPatho = dfPatho.transpose()
  dfPatho.rename(columns={0: 'Count'},inplace=True)
  dfPatho = dfPatho.sort_values(by=['Count'],ascending=False)
  a = np.array(dfPatho['Count'])
  sum = np.sum(a,axis=0)
  a = np.divide(a,sum)
  a = np.multiply(a,100)
  a = np.around(a, 2)
  dfPatho['Ratio'] = a.tolist()
  display(HTML("<h2> Number of pathologies : "+str(nbPatho)+"</h2>"))
  display(HTML("<h2> Total : "+str(sum)+"</h2>"))
  return dfPatho

dfHosp = computeFreq("Hospitalisation")
data_table.DataTable(dfHosp)
```

```
<IPython.core.display.HTML object>


<IPython.core.display.HTML object>
```

[153]:
```
<google.colab.data_table.DataTable object>
```

Liste des pathologies ayant entraîné le plus de décès suite à la vaccination Afin d'extraire les pathologies fatales qui ont entraînées de décès et qui émergent des données, nous calculons les fréquences d'apparition de ces pathologies

[154]:
```
def computeFreqFatal(search):
  df = dict()
  reactions = df0['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)']
  for index, value in reactions.items():
    reaction = str(value)
    reaction = reaction.split("<BR><BR>")
    for r in reaction :
      if r.find(search)!=-1:
        r = parseReaction(r)
        for re in r :
          if re!="" and re!=" ":
            if search=="Fatal":
              if index not in df.keys():
                df.update({index:[1]})
```

```
dfFatal = computeFreq("Fatal")
data_table.DataTable(dfFatal)
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

[154]:
```
<google.colab.data_table.DataTable object>
```

Nous pouvons remarquer que les "causes inconnues" sont celles qui entraînent une très grande partie des décès dû au vaccin.

Taux d'hospitalisation et de létalité dû au vaccin en cas d'EI par sexe

[155]:
```
from google.colab import data_table

features = {"Patient Sex" : ["Female","Male"],
            "Patient Age Group":["0-1 Month","2 Months - 2 Years", "3-11␣
↪Years","12-17 Years", "18-64 Years", "65-85 Years","More than 85 Years"]}

def computeFreqSex(sex, search):
  dfHosp = dict()
  reactions = df0[['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)',"Patient Sex"]]
  reactions = reactions[reactions["Patient Sex"]==sex]
  reactions = reactions['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)']
  count = len(df0[df0["Patient Sex"]==sex])
  totalCount = count + len(df1[df1["Patient Sex"]==sex])
  nb = 0
  for index, value in reactions.items():
    reaction = str(value)
    reaction = reaction.split("<BR><BR>")
    for r in reaction :
      if r.find(search)!=-1:
        nb+=1
        break
  if totalCount!=0 : x1 = round((nb/totalCount)*100,2)
  else : x1 = None
  if count!=0 : x2 = round((nb/count)*100,2)
  else : x2 = None
  return x1, x2

resultBySex = pandas.DataFrame(columns=['Patient Sex', 'Taux hospitalisation␣
↪(si effet indesirable)', 'Taux hospitalisation (si effet indésirable grave)',
```

```
                                    'Létalité du vaccin (si effet indésirable)',␣
↪'Létalité du vaccin (si effet indésirable grave)'])
i=0
for item in features["Patient Sex"]:
  ratio1, ratio2 = computeFreqSex(item,"Hospitalisation")
  ratioDeath1, ratioDeath2 = computeFreqSex(item,"Fatal")
  resultBySex.loc[i]=[item,ratio1,ratio2,ratioDeath1,ratioDeath2]
  i+=1
data_table.DataTable(resultBySex,include_index=False)
```

[155]: <google.colab.data_table.DataTable object>

Taux d'hospitalisation et de létalité dû au vaccin en cas d'EI par classe d'âge

[156]:
```
def computeFreqAge(age, search):
  dfHosp = dict()
  reactions = df0[['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)',"Patient Age Group"]]
  reactions = reactions[reactions["Patient Age Group"]==age]
  reactions = reactions['Reaction List PT (Duration - Outcome - Seriousness␣
↪Criteria)']
  count = len(df0[df0["Patient Age Group"]==age])
  totalCount = count + len(df1[df1["Patient Age Group"]==age])
  nb = 0
  for index, value in reactions.items():
    reaction = str(value)
    reaction = reaction.split("<BR><BR>")
    for r in reaction :
      if r.find(search)!=-1:
        nb+=1
        break
  if totalCount!=0 : x1 = round((nb/totalCount)*100,2)
  else : x1 = None
  if count!=0 : x2 = round((nb/count)*100,2)
  else : x2 = None
  return x1, x2

resultByAgeGroup = pandas.DataFrame(columns=['Patient Sex', 'Taux␣
↪hospitalisation (si effet indesirable)', 'Taux hospitalisation (si effet␣
↪indésirable grave)',
                                    'Létalité du vaccin (si effet indésirable)',␣
↪'Létalité du vaccin (si effet indésirable grave)'])
i=0
for item in features["Patient Age Group"]:
  ratio1, ratio2 = computeFreqAge(item,"Hospitalisation")
  ratioDeath1, ratioDeath2 = computeFreqAge(item,"Fatal")
```

```
  resultByAgeGroup.loc[i]=[item,ratio1,ratio2,ratioDeath1,ratioDeath2]

  i+=1
data_table.DataTable(resultByAgeGroup,include_index=False)
```

[156]: <google.colab.data_table.DataTable object>

4.5 Calcul des GLM: prédiction et OR

Sévérité vs sexe Dans ce cas, on a le risque retenu est celui de la sévérité (variabel Y_1) tandis que les facteur est celui du sexe (variable X_1). On cherche donc le GLM associé à

$$\texttt{logit}\left(P\left(Y_1 = S \mid X_1\right)\right) = \beta_0 + \beta_1 X_1^1 \tag{46}$$

où la variable X_1^1 est choisie telle que

$$X_1^1\left(\omega\right) = 0 \;\Leftrightarrow X_1\left(\omega\right) = F \tag{47}$$

$$X_1^1\left(\omega\right) = 1 \;\Leftrightarrow X_1\left(\omega\right) = M \tag{48}$$

[157]:
```
import statsmodels.api as sm
import statsmodels.formula.api as smf
import scipy
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.display import display, HTML
import warnings
warnings.filterwarnings('ignore')

ref_age = ["0-1 Month", "2 Months - 2 Years", "3-11 Years","12-17 Years",␣
 ↪"18-64 Years", "65-85 Years","More than 85 Years"]

def addORPerSex(odds_ratios,result,label):
  result.update({label+" H vs F":[odds_ratios.iloc[1,0],odds_ratios.
 ↪iloc[1,1],odds_ratios.iloc[1,2]]})
  return result

def addORPerAge(odds_ratios,result,label) :
  result.update({label+" 0-1M vs 65-85years":[odds_ratios.iloc[1,0],odds_ratios.
 ↪iloc[1,1],odds_ratios.iloc[1,2]]})
  result.update({label+" 2M-2Y vs 65-85years":[odds_ratios.
 ↪iloc[2,0],odds_ratios.iloc[2,1],odds_ratios.iloc[2,2]]})
```

```
  result.update({label+" 3-11Y vs 65-85years":[odds_ratios.
↪iloc[3,0],odds_ratios.iloc[3,1],odds_ratios.iloc[3,2]]})
  result.update({label+" 12-17Y vs 65-85years":[odds_ratios.
↪iloc[4,0],odds_ratios.iloc[4,1],odds_ratios.iloc[4,2]]})
  result.update({label+" 18-64Y vs 65-85years":[odds_ratios.
↪iloc[5,0],odds_ratios.iloc[5,1],odds_ratios.iloc[5,2]]})
  result.update({label+" +85Y vs 65-85years":[odds_ratios.iloc[6,0],odds_ratios.
↪iloc[6,1],odds_ratios.iloc[6,2]]})
  return result

def computeModel(test, df2):
  glmresults = smf.glm(formula="Severity ~ "+test, data=df2, family=sm.families.
↪Binomial()).fit()
  param = glmresults.params
  CI = glmresults.conf_int(alpha = 0.05)
  pval = glmresults.pvalues
  glmresults.summary()
  odds_ratios = pandas.DataFrame(
    {
        "OR": glmresults.params,
        "Lower CI": glmresults.conf_int()[0],
        "Upper CI": glmresults.conf_int()[1],
    }
  )
  odds_ratios = np.exp(odds_ratios)
  return param, odds_ratios

def probasPerAge(params):
  ageModel = []
  probas = []
  for i in range(0,7):
    age = ref_age[i]
    if i<6 and i!=5 :
      p = np.round(scipy.special.expit(params[i+1] + params[0]),2)
    elif i==5 :
      p = np.round(scipy.special.expit(params[0]),2)
    elif i==6:
      p = np.round(scipy.special.expit(params[i] + params[0]),2)
    probas.append(p*100)
    ageModel.append(age)
  return ageModel, probas

def plotResultBySex(title,nameFile,params):
  sns.reset_defaults()
  x = ["Femmes","Hommes"]
  y = [np.round(scipy.special.expit(params[0]),2)*100,np.round(scipy.special.
↪expit(params[0] + params[1]),2)*100]
```

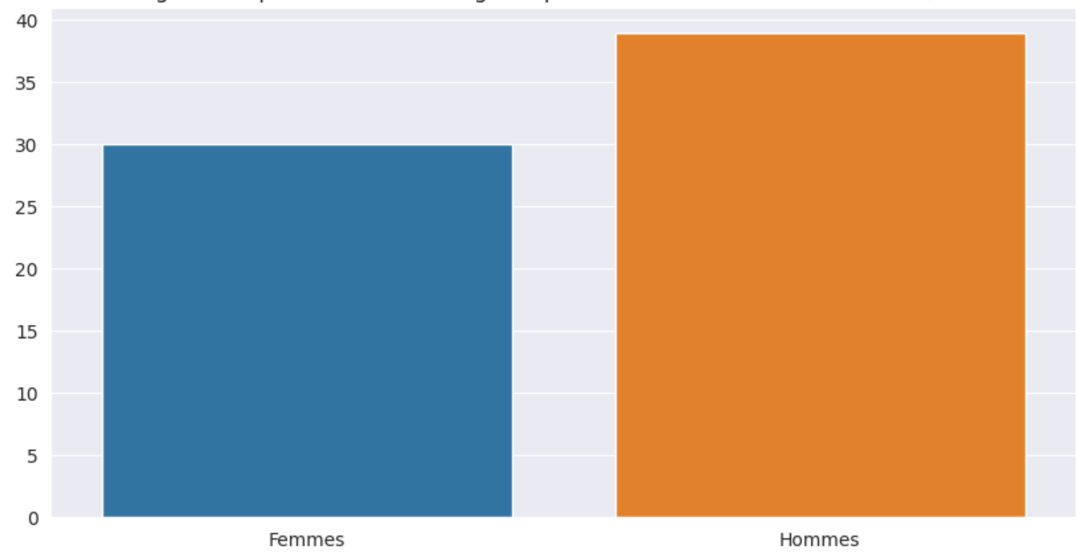
```
  plt.figure(figsize=(10,5))
  sns.set_style('darkgrid')
  sns.barplot(x, y)
  plt.title("Pourcentage de risque d'avoir un effet grave par sexe en cas␣
↪d'effet indésirable / vaccin "+vaccin)
  plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/pcMFEIS_"+vaccin+".png")
  plt.show()

def plotResultByAge(title, nameFile, ageModel, probas):
  plt.figure(figsize=(10,5))
  sns.set_style('darkgrid')
  sns.barplot(ageModel, probas)
  plt.title(title)
  plt.savefig(nameFile)
  plt.show()

result = dict()
df2.rename(columns = {"Patient Sex": "Sex"}, inplace = True)
df2.rename(columns = {"Patient Age Group": "PAG"}, inplace = True)
df2 = df2.replace({"Years 65-85": "- Years 65-85"})
params, odds_ratio = computeModel("Sex",df2)
result = addORPerSex(odds_ratio,result,"Severity")

plotResultBySex("Pourcentage de risque d'avoir un effet grave par sexe en cas␣
↪d'effet indésirable / vaccin "+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/pcMFEIS_"+vaccin+".
↪png",
                params)
```





Sévérité vs Classe d'âge Le risque retenu est toujours la variable Y_1 (c'est à dire la sévérité), par contre le facteur est celui de la classe d'âge (varibale X_2). Cette variable prend les modalités suivantes

$$X_2 \in \{I_0, \cdots, I_6\} \tag{49}$$
$$I_0 = \textit{ Month 0-1, etc.} \tag{50}$$

La variable X_2 prend 7 modalités (on laisse tomber évidemment les cas où la classe d'âge n'est pas renseignée). Pour le calcul des Odds ratios, on s'arrange pour que l'intervalle $I_3 =$ *Years 12-17* (l'idée ici c'est de s'intéresser aux effets sur les enfants de la vaccination) soit choisi comme référence dans le GLM. A partir de la variable X_2 on va donc créer les 6 variables suivantes

$$X_2^1(\omega) = 1 \Leftrightarrow X_2(\omega) = I_0, \textit{ 0 sinon} \tag{51}$$
$$X_2^2(\omega) = 1 \Leftrightarrow X_2(\omega) = I_1, \textit{ 0 sinon} \tag{52}$$
$$X_2^3(\omega) = 1 \Leftrightarrow X_2(\omega) = I_2, \textit{ 0 sinon} \tag{53}$$
$$X_2^4(\omega) = 1 \Leftrightarrow X_2(\omega) = I_4, \textit{ 0 sinon} \tag{54}$$
$$X_2^5(\omega) = 1 \Leftrightarrow X_2(\omega) = I_5, \textit{ 0 sinon} \tag{55}$$
$$X_2^6(\omega) = 1 \Leftrightarrow X_2(\omega) = I_6, \textit{ 0 sinon} \tag{56}$$

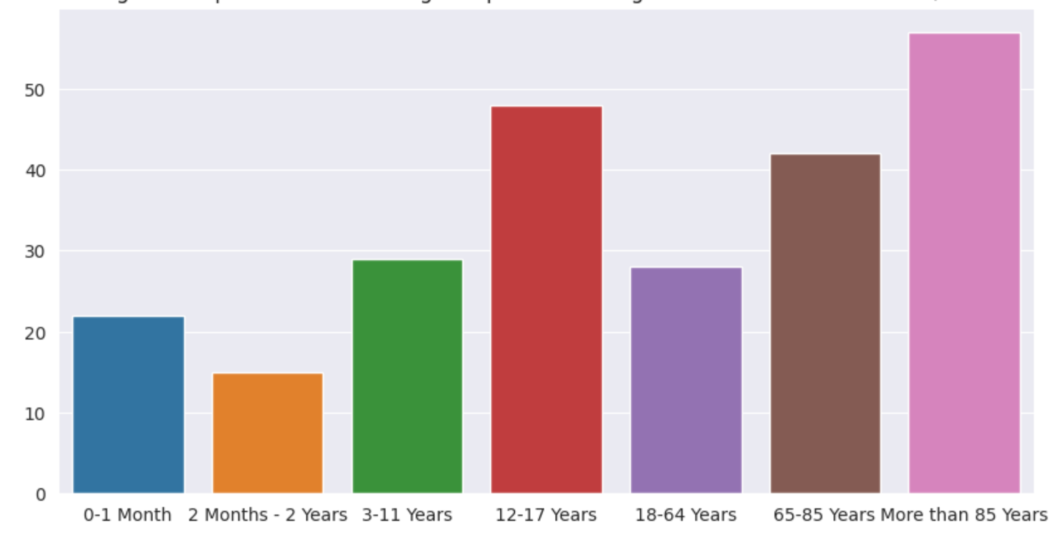
On écrit ensuite que l'on a

$$\texttt{logit}\left(P\left(Y_1 = S \mid X_2\right)\right) = \beta_0 + \sum_{j=1}^{j=6} \beta_j X_2^j \tag{57}$$

Les valeurs des coefficients β sont obtenus par le GLM

[158]:
```
params, odds_ratios = computeModel("PAG",df2)
result = addORPerAge(odds_ratios,result,"Severity")
ageModel, probas = probasPerAge(params)
plotResultByAge("Pourcentage de risque d'avoir un effet grave par classe d'âge␣
 ↪en cas d'effet indésirable / vaccin "+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/pcEIAGE_"+vaccin+".
 ↪png",
                ageModel, probas)
```





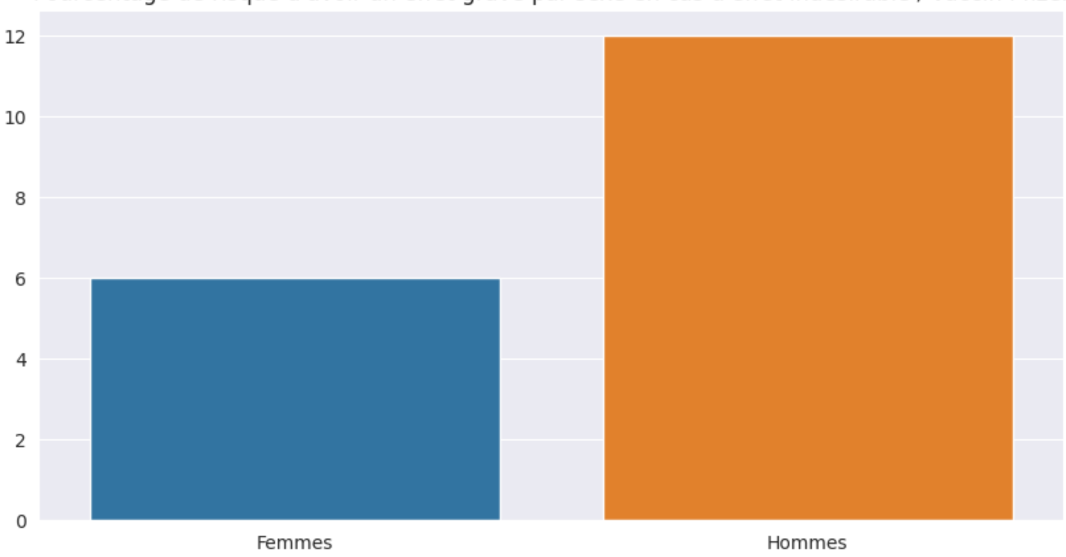
Ainsi, parmi les personnes qui ont des effets indésirables, la classe d'âge des $12-17$ expérimente des effets graves de manière significativement supérieure à n'importe quelle autre classe d'âge. Cela est totalement contradictoire avec le fait que les formes graves du covid, lorsqu'il est symptomatique, épargne totalement cette classe d'âge.

Hospitalisation vs sexe

[159]:
```
def searchFeature(df, search, value1, value2):
  values = []
  reactions = df["Reaction List PT (Duration - Outcome - Seriousness Criteria)"]
  for index, value in reactions.items():
    reaction = str(value)
    reaction = reaction.split("<BR><BR>")
    trouve = False
    for r in reaction :
      if r.find(search)!=-1:
        values.append(value1)
        trouve = True
        break
    if not trouve : values.append(value2)
  df["Severity"] = values
  return df

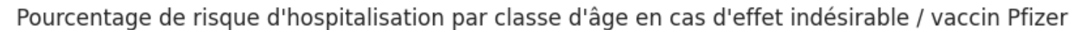
df3 = df2.copy()
df3 = df3.replace({"Years 65-85": "- Years 65-85"})
df3["Severity"]=False
df3 = searchFeature(df3,"Hospitalisation","Hospitalisé","Non hospitalisé")
params, odds_ratios = computeModel("Sex",df3)
```

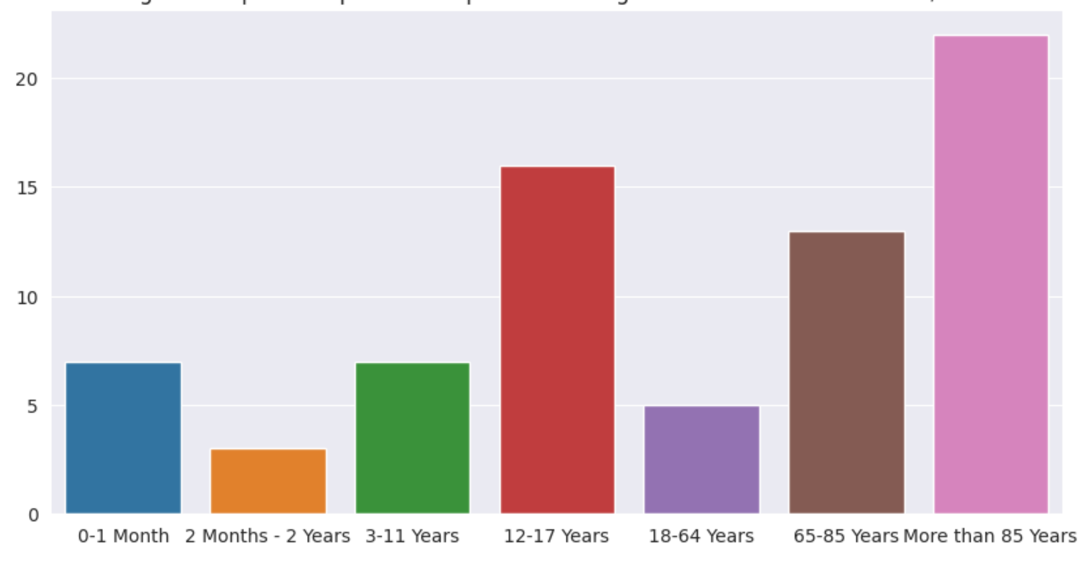
```
result = addORPerSex(odds_ratios,result,"Hospitalisation")
plotResultBySex("Pourcentage de risque d'hospitalisation par sexe en cas␣
↪d'effet indésirable / vaccin "+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/pcHospSexe_"+vaccin+".
↪png",
                params)
```



Hospitalisation par classe d'âge

[160]:
```
params, odds_ratios = computeModel('PAG',df3)
result = addORPerAge(odds_ratios,result,"Hospitalisation")
ageModel, probas = probasPerAge(params)
plotResultByAge("Pourcentage de risque d'hospitalisation par classe d'âge en␣
↪cas d'effet indésirable / vaccin "+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/pcHosAge_"+vaccin+".
↪png",
                ageModel, probas)
```

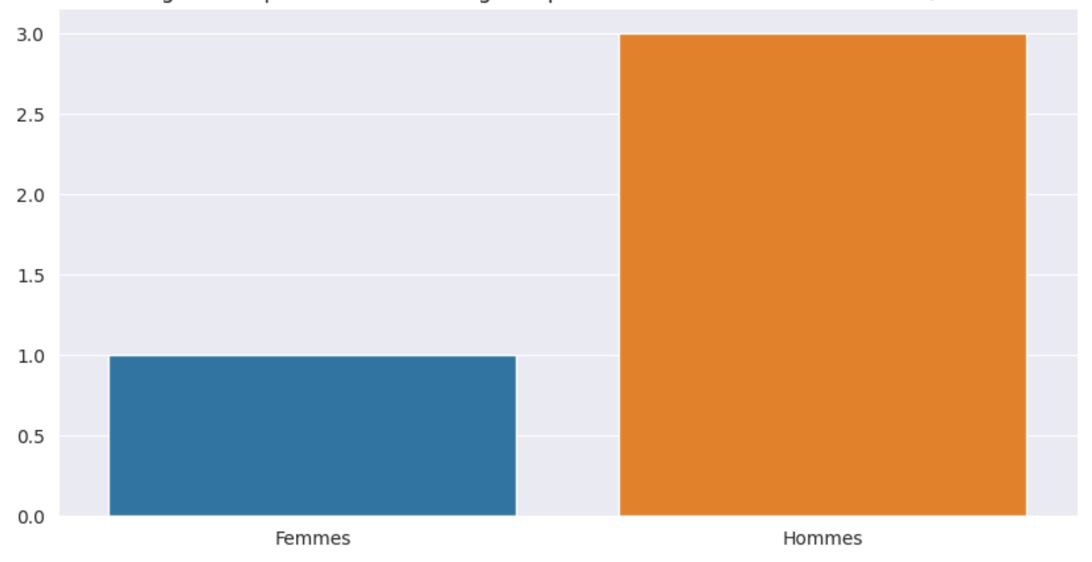




Mort vs sexe Outre les formes graves, un autre indicateur important est la probabilité de décès en fonction des facteurs. De ce point de vue, nous choisissons maintenant le risque Y_2 avec la variable explicative X_1. Pour mettre en place les GLM, il faut d'abord créer un dataframe dans lequel il y a une variable aléatoire renvoyant la fatalité ou non du vaccin. Dans les données européennes EudraVigilance telles qu'elles sont disponibles, cette variable n'est pas codée directement. Il faut donc créer d'abord le dataframe adequat.

```
[161]: df3 = df2.copy()
df3["Severity"]=False
df3 = searchFeature(df3,"Fatal","Mort","Vivant")
params, odds_ratio = computeModel("Sex", df3)
result = addORPerSex(odds_ratios,result,"Death")
plotResultBySex("Pourcentage de risque de décès par sexe en cas d'EI / vaccin␣
↪"+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/
↪letaliteSexe_"+vaccin+".png",
                params)
```

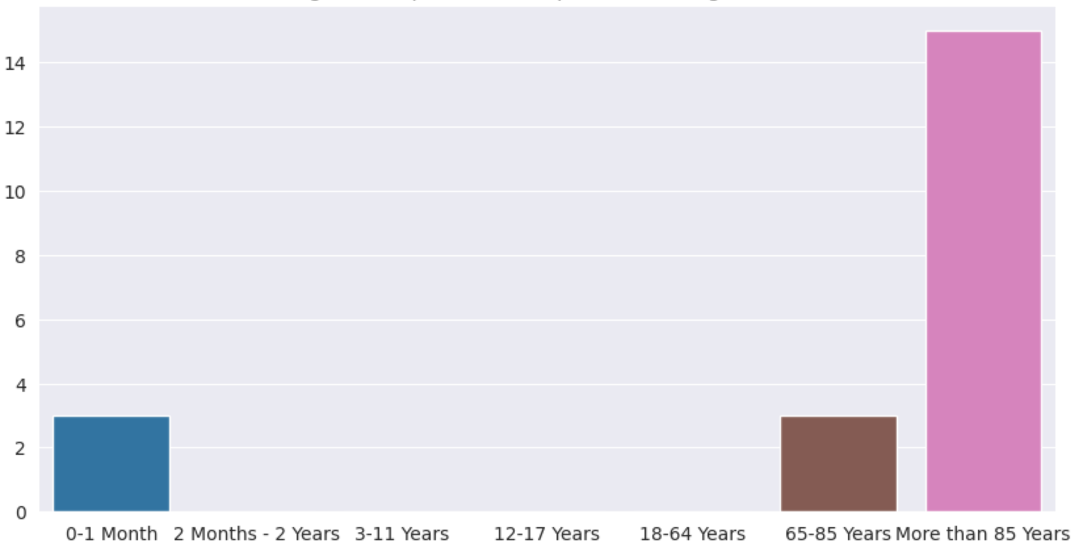




Mort vs Classe d'âge

[162]:
```
params, odds_ratios = computeModel('PAG',df3)
result = addORPerAge(odds_ratios,result,"Death")
ageModel,probas=probasPerAge(params)
plotResultByAge("Pourcentage de risque de décès par classe d'âge / vaccin␣
 ↪"+vaccin,
                "/content/drive/MyDrive/MetaAnalyseVaccin/
 ↪letaliteAGE_"+vaccin+".png",
                ageModel, probas)
```



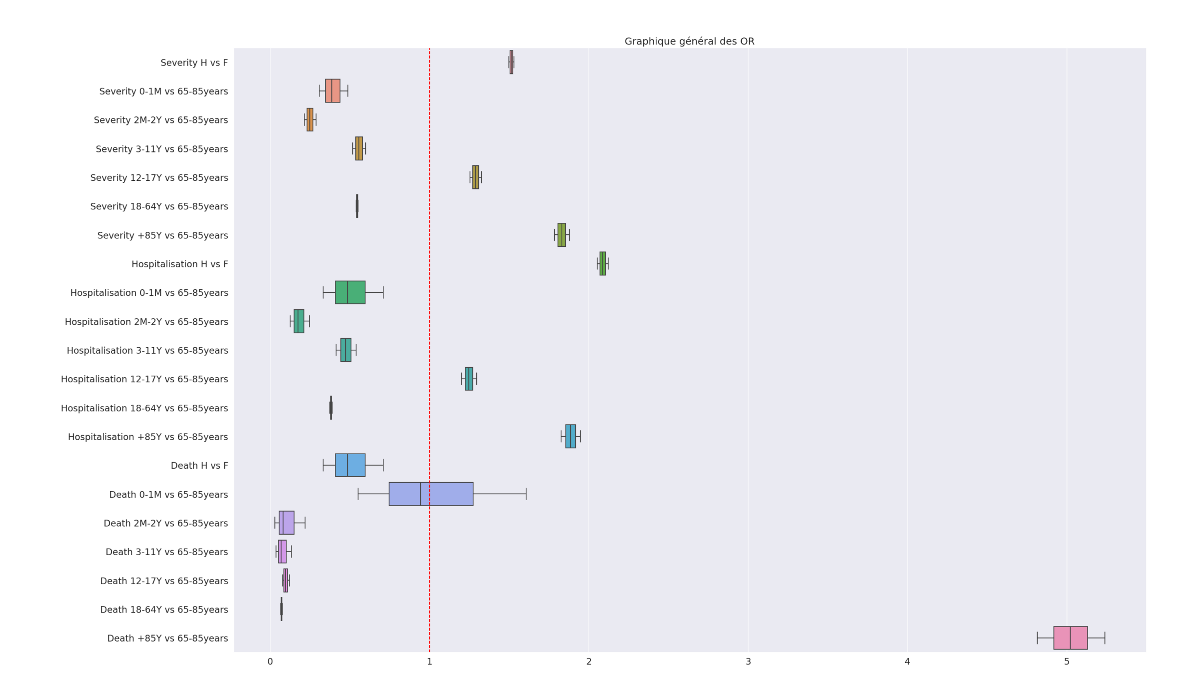


Graphique général des OR

[163]:
```
dfResult = pandas.DataFrame.from_dict(result)
dfResult = dfResult.transpose()
dfResult.rename(columns={0: 'OR', 1:"lci", 2:"uci"},inplace=True)
dfResult.to_excel(path_file_data+'tableResult.
 ↪xlsx',sheet_name='sheet1',index=True)
data_table.DataTable(dfResult)
```

[163]: `<google.colab.data_table.DataTable object>`

[164]:
```
sns.set_theme(style="darkgrid")
sns.set(font_scale=1.5)
plt.figure(figsize=(30,20))
dfResult = dfResult.T
sns.boxplot(data=dfResult, orient="h", showfliers = False)
plt.axvline(1.0, color='red', linestyle='dashed')
plt.title("Graphique général des OR")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/OR_"+vaccin+".png")
plt.show()
dfResult = dfResult.T
sns.reset_defaults()
```



4.6 Comparaison avec le Covid19 (calcul de la balance bénéfices / risques)

Les données receuillies sont issues de la base de données COVerAGE-DB https://osf.io/mpwjq/wiki/home/ qui recence les cas / décès par age, sexe et pays : https://osf.io/9dsfk/

Nous nous limiterons ici à la France afin de calculer les facteurs de risque de décès par sexe et classe.

Pour pouvoir comparer ces facteurs de risque par classes d'âge aux facteurs de riques vaccinaux précédemment calculés, nous utiliserons les classes d'âge des données européennes

Extraction des données cas / décès par sexe et age pour la France

[113]:
```
idData = "1nxp6aNwH8oI3KEU65_z_ro6e3M8F0XaL"
nameFileData = "Data/inputDB.csv"
downloaded = drive.CreateFile({"id":idData})
downloaded.GetContentFile(path_file_data+nameFileData)

import pandas
osf = pandas.read_csv("/content/drive/MyDrive/MetaAnalyseVaccin/datas/Data/
↪inputDB.csv", sep=",", encoding='latin-1')
new_header = osf.iloc[0]
#new_header = new_header[0].split(",")
osf = osf[1:]
osf.columns = new_header
```

```
osf.to_csv("/content/drive/MyDrive/MetaAnalyseVaccin/datas/Data/datas.csv")
osf = pandas.read_csv("/content/drive/MyDrive/MetaAnalyseVaccin/datas/Data/
↪datas.csv")

osf.rename(columns = {"Unnamed: 0": "Country"}, inplace = True)
osf.rename(columns = {"Unnamed: 1": "Region"}, inplace = True)
osf.rename(columns = {"Unnamed: 2": "Date"}, inplace = True)
osf.rename(columns = {"Unnamed: 3": "DateInt"}, inplace = True)
osf.rename(columns = {"Unnamed: 4": "Sex"}, inplace = True)
osf.rename(columns = {"Unnamed: 5": "Age"}, inplace = True)
osf.rename(columns = {"Unnamed: 6": "AgeInt"}, inplace = True)
osf.rename(columns = {"Unnamed: 7": "Comptage"}, inplace = True)
osf.rename(columns = {"Unnamed: 8": "Measure"}, inplace = True)
osf.rename(columns = {"Unnamed: 9": "Cumulative"}, inplace = True)
osf.rename(columns = {"Unnamed: 10": "codeRegion"}, inplace = True)
```

[130]:
```
#Données pour la France
osf_france = osf[(osf["Measure"]=="Cases")&(osf["Country"]=="France")&␣
↪(osf["Region"]=="All")]
osf_france = osf[(osf["DateInt"]=="14.02.2022") & (osf["Country"]=="France") &␣
↪(osf["Region"]=="All")]
cases = osf_france[osf_france["Measure"]=="Cases"]
data_table.DataTable(cases)
```

[130]: <google.colab.data_table.DataTable object>

[115]:
```
deaths = osf_france[osf_france["Measure"]=="Deaths"]
data_table.DataTable(deaths)
```

[115]: <google.colab.data_table.DataTable object>

4.6.1 Calcul du taux de létalité apparent du Covid19 par sexe et par classe d'âge

Calcul de létalité apparente par sexe

[128]:
```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

def letalitySexRateCountry(codeCountry):
  osf_country = osf[(osf["Country"]==codeCountry) & (osf["Region"]=="All")]
  if len(osf_country)>0 :
    osf_country = osf_country.sort_values(by = 'DateInt', ascending = False)
    date = osf_country.iloc[0]['DateInt']
    osf_country = osf_country[osf_country["DateInt"]==date]
```

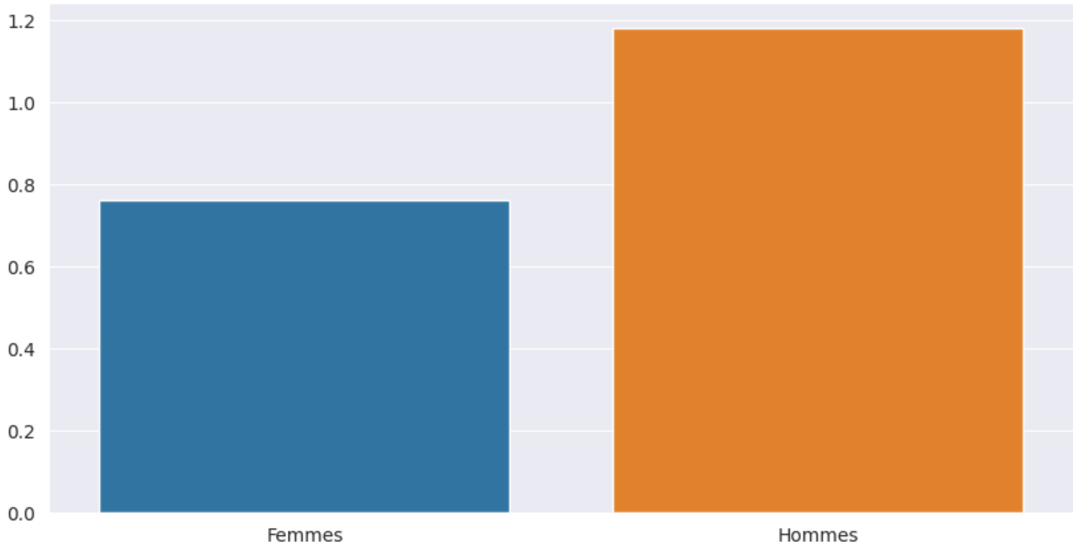
```
    cases = osf_country[osf_country["Measure"]=="Cases"]
    deaths = osf_country[osf_country["Measure"]=="Deaths"]
    nbCasesH = np.sum(np.array(cases[(cases['Sex']=='m') & (cases["Age"]!
↪="TOT")]["Cumulative"]),axis=0)
    nbCasesF = np.sum(np.array(cases[(cases['Sex']=='f') & (cases["Age"]!
↪="TOT")]["Cumulative"]),axis=0)
    nbDeathsH = np.sum(np.array(deaths[(deaths['Sex']=='m') & (deaths["Age"]!
↪="TOT")]["Cumulative"]),axis=0)
    nbDeathsF = np.sum(np.array(deaths[(deaths['Sex']=='f') & (deaths["Age"]!
↪="TOT")]["Cumulative"]),axis=0)
    # Si pas de cumul => dans les données, cela correspond à la fraction sur␣
↪les femmes et il est impossible de calculer
    # le taux de létalité car nous n'avons pas le nombre de mort mais juste un␣
↪ratio H/F
    if nbCasesH==0 or nbCasesF==0 or nbDeathsH==0 or nbDeathsF==0 :
      return 0,0

    rateH = (nbDeathsH/nbCasesH)*100
    rateF = (nbDeathsF/nbCasesF)*100
    return rateH, rateF
  else : return 0,0

#Exemple de calcul de la létalité apparente du Covid 19 par sexe
rateCovidHomme, rateCovidFemme = letalitySexRateCountry("France")
print("Létalité apparente (%) pour les hommes :",rateCovidHomme)
print("Létalité apparente (%) pour les femmes :",rateCovidFemme)
sns.reset_defaults()
x = ["Femmes","Hommes"]
y = [rateCovidFemme,rateCovidHomme]
plt.figure(figsize=(10,5))
sns.set_style('darkgrid')
sns.barplot(x, y)
plt.title("Létalité apparente (%) du Covid19 par sexe pour la France")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/letaliteCOVSexe.png")
plt.show()
```

```
Létalité apparente (%) pour les hommes : 1.180588428455066
Létalité apparente (%) pour les femmes : 0.759698901124776
```





Malheureusement, il n'est pas toujours possible de comparer ce taux par sexe pour tous les pays européens. En effet, ne recense pas toujours les décès.

Par exemple, pour la croatie, les dernières données en date remontent au 12.03.2020 et seul les cas ont été remontés dans ces données et non les décès.

[117]:
```
osf_country = osf[(osf["Country"]=="Croatia") & (osf["Region"]=="All")]
osf_country = osf_country.sort_values(by = 'DateInt', ascending = False)
date = osf_country.iloc[0]['DateInt']
osf_country = osf_country[osf_country["DateInt"]==date]
cases = osf_country[osf_country["Measure"]=="Cases"]
deaths = osf_country[osf_country["Measure"]=="Deaths"]
data_table.DataTable(osf_country)
```

[117]: <google.colab.data_table.DataTable object>

Aussi nous prendrons en compte les taux de létalités des pays qui font la distinction des cas ou des décès selon le sexe et qui ont bien recensé le nombre de cas et de décès par sexe.

[129]:
```
import math
listCountry=["Germany",␣
↪"Austria","Belgium","Bulgaria","Cyprus","Croatia","Denmark","Spain","Estonia","Finland",
           ␣
↪"France","Greece","Hungary","Italy","Ireland","Latvia","Lithuania","Luxembourg","Malta","th
↪Netherlands",
             "Poland","Portugal","the Czech␣
↪Republic","Romania","Slovakia","Slovenia","Sweden"]
```

```
letalityCovidSex = pandas.DataFrame()
ratesH = []
ratesF = []
countries = []
for country in listCountry :
  rateH, rateF = letalitySexRateCountry(country)
  if rateH>0 and rateF>0 and not math.isinf(rateH) and not math.isinf(rateF):
    ratesH.append(rateH)
    ratesF.append(rateF)
    countries.append(country)

letalityCovidSex["Country"]=countries
letalityCovidSex["Femmes"]=ratesF
letalityCovidSex["Hommes"]=ratesH

avgH = np.average(np.array(letalityCovidSex["Hommes"]))
avgF = np.average(np.array(letalityCovidSex["Femmes"]))

data_table.DataTable(letalityCovidSex)
```

[129]: <google.colab.data_table.DataTable object>

Calcul du taux de létalité apparente par classe d'âge Les classes d'âge dans les données sur le Covid19 sont différentes de celles des données concernant le vaccin. Celle-ci sont également propres à chacun des pays.

A titre d'exemple, dans les données, la dernière classe d'âge pour l'Allemagne correspond aux 80 ans et plus !

Avec les données présentes, il n'est pas possible pour le moment d'établir ce calcul de façon uniforme pour chaque classe d'âge pour l'ensemble des pays européens.

Aussi, nous calculerons la létalité apparente du Covid19 en France sur les classes d'âge suivante : - 0-9 ans / pour comparer sur la moyenne des létalités des classes d'age inférieures à 11 ans pour le vaccin - **10-19 ans** / pour comparer sur la classe d'âge 12-17 ans du vaccin - **20-69 ans** / pour comparer sur la classe d'âge 18-65 ans du vaccin - **70-89 ans** / pour comparer sur la classe d'âge 65-85 ans du vaccin - **90 et plus** / pour comparer sur la classe d'âge +85 ans du vaccin

[132]:
```
classes = [0,10,20,70,90]
ca = []
rates = []
def letalityAge(codeCountry):
  osf_country = osf[(osf["Country"]==codeCountry) & (osf["Region"]=="All")]
  if len(osf_country)>0 :
    ages = []
    rates = []
    result = pandas.DataFrame()
    osf_country = osf_country.sort_values(by = 'DateInt', ascending = False)
```

```
    date = osf_country.iloc[0]['DateInt']
    osf_country = osf_country[(osf_country["DateInt"]==date)]
    osf_country.drop(osf_country[(osf_country["AgeInt"] == "UNK") |␣
↪(osf_country["Age"] == "TOT") | (osf_country["Age"] == "UNK")].index,␣
↪inplace=True)
    osf_country["Age"]= osf_country['Age'].astype(int)
    max_age = np.max(np.array(osf_country["Age"]))
    print(max_age)
    for i in range(0,len(classes)-1):
      cases = osf_country[(osf_country["Measure"]=="Cases") &␣
↪(osf_country['Sex']=='b') & (osf_country['Age']<classes[i+1]) &␣
↪(osf_country['Age']>=classes[i])]
      deaths = osf_country[(osf_country["Measure"]=="Deaths") &␣
↪(osf_country['Sex']=='b') & (osf_country['Age']<classes[i+1]) &␣
↪(osf_country['Age']>=classes[i])]
      # La valeur 'b' n'existe pas pour le pays
      if len(cases)==0 :
        cases = osf_country[(osf_country["Measure"]=="Cases") &␣
↪((osf_country['Sex']=='m') | (osf_country['Sex']=='f')) &␣
↪(osf_country['Age']<classes[i+1]) & (osf_country['Age']>=classes[i])]
        deaths = osf_country[(osf_country["Measure"]=="Deaths") &␣
↪(osf_country['Age']<classes[i+1]) & (osf_country['Age']>=classes[i])]

      nbCases = np.sum(np.array(cases["Cumulative"]))
      nbDeaths = np.sum(np.array(deaths["Cumulative"]))
      rate = (nbDeaths/nbCases)*100
      age = str(classes[i])+"-"+str(classes[i+1]-1)
      ages.append(age)
      rates.append(rate)
    # Cas de la dernière classe d'âge
    cases = osf_country[(osf_country["Measure"]=="Cases") &␣
↪(osf_country['Sex']=='b') & (osf_country['Age']>=classes[-1])]
    deaths = osf_country[(osf_country["Measure"]=="Deaths") &␣
↪(osf_country['Sex']=='b') & (osf_country['Age']>=classes[-1])]
    # La valeur 'b' n'existe pas pour le pays
    if len(cases)==0 :
      cases = osf_country[(osf_country["Measure"]=="Cases") &␣
↪((osf_country['Sex']=='m') | (osf_country['Sex']=='f')) &␣
↪(osf_country['Age']>=classes[-1])]
      deaths = osf_country[(osf_country["Measure"]=="Deaths") &␣
↪((osf_country['Sex']=='m') | (osf_country['Sex']=='f')) &␣
↪(osf_country['Age']>=classes[-1])]

    deaths = osf_country[(osf_country["Measure"]=="Deaths") &␣
↪(osf_country['Sex']=='b') & (osf_country['Age']>=classes[-1])]
    nbCases = np.sum(np.array(cases["Cumulative"]))
```

```
    nbDeaths = np.sum(np.array(deaths["Cumulative"]))

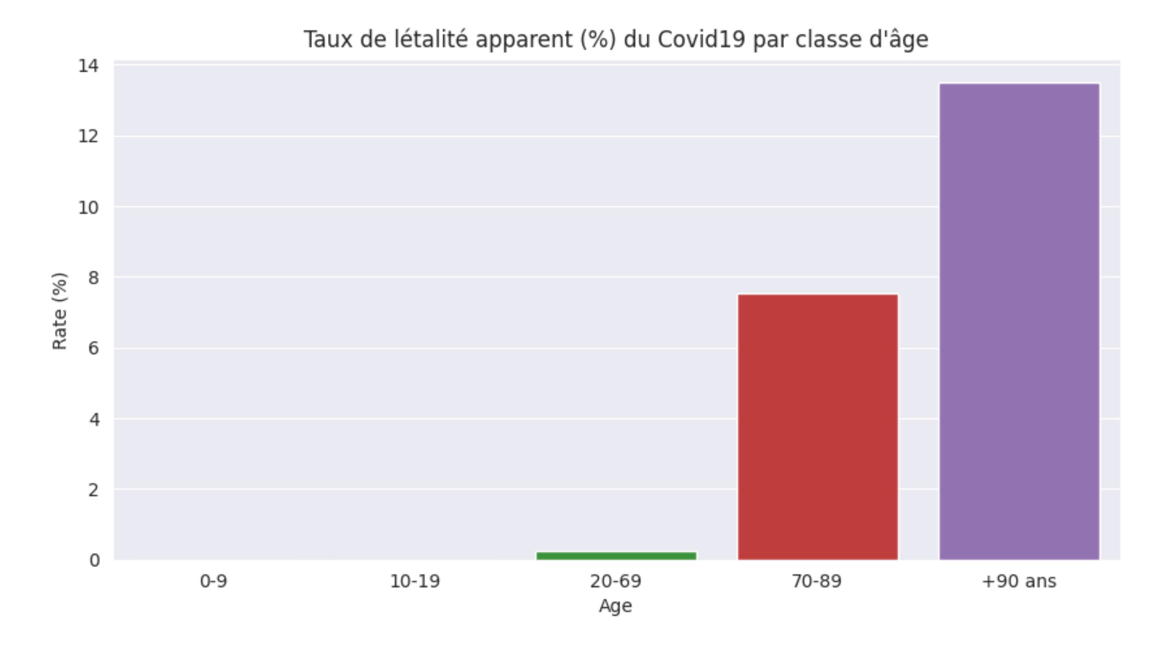
    rate = (nbDeaths/nbCases)*100
    age = "+90 ans"
    ages.append(age)
    rates.append(rate)
    result["Age"]=ages
    result["Rate (%)"]=rates
    return result

resultCovidAge = letalityAge("France")
data_table.DataTable(resultCovidAge)
```

90

[132]: <google.colab.data_table.DataTable object>

```
[133]: plt.figure(figsize=(10,5))
sns.set_style('darkgrid')
sns.barplot(resultCovidAge["Age"],resultCovidAge["Rate (%)"])
plt.title("Taux de létalité apparent (%) du Covid19 par classe d'âge")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/letaliteCOVAge.png")
plt.show()
```

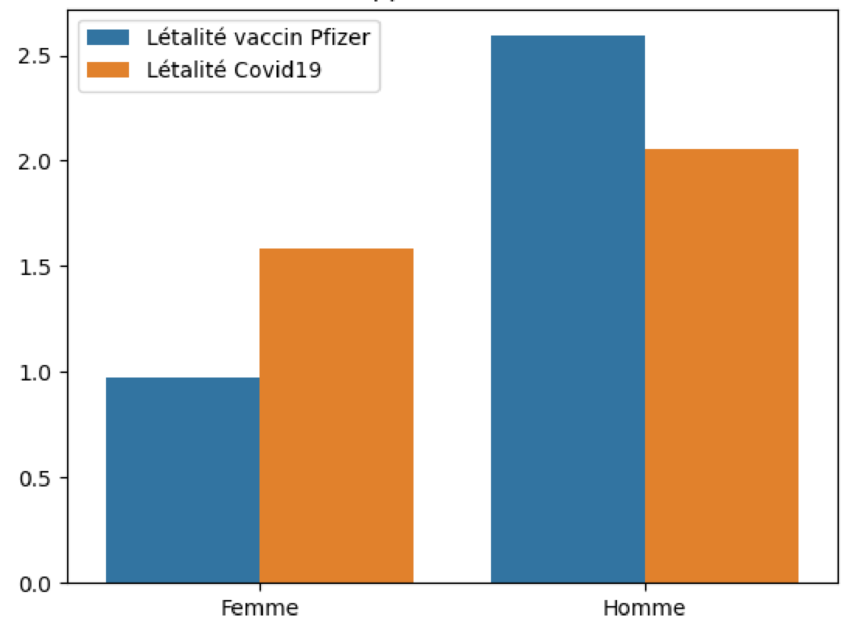


5 Balance bénéfice risque

Nous présentons ici la balance bénéfice risque au regard de la létalité apparente du vaccin selectionné et celle du Covid19 : - par sexe - par classe d'âges

5.0.1 Balance par rapport au sexe

[165]:
```
raw_data = {
    # cat: Vaccin              Covid19
    'x': ['Femme',            'Femme',
          'Homme',            'Homme'],
    'y': [resultBySex['Létalité du vaccin (si effet indésirable)'][0], avgF,
          resultBySex['Létalité du vaccin (si effet indésirable)'][1], avgH ],
    'category': ['Létalité vaccin '+vaccin, 'Létalité Covid19', 'Létalité␣
 ↪vaccin '+vaccin, 'Létalité Covid19']
           }
sns.barplot(x='x', y='y', hue='category', data=raw_data)
sns.set_style('darkgrid')
plt.title("Comparaison des taux de létalité apparent (%) entre le vaccin␣
 ↪"+vaccin+" et le Covid19")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/
 ↪completalitesexe_"+vaccin+".png")
plt.show()
```



5.0.2 Balance par rapport sur les classes d'âge entre la létalité apparente du Covid19 en France et celle du vaccin

[166]:
```
avg_first_ageGroup = np.average(np.array(probasDeathAge[0:3]))

compare = pandas.DataFrame()
ages = ['[0-11]\n[0-9] ans', '[12-17]\n[10-20] ans', '[18-65]\n[20-69]␣
↪ans','[65-85]\n[70-89] ans', '[+85]\n[+90] ans']
lv = resultByAgeGroup['Létalité du vaccin (si effet indésirable)']
avgFirst = 0
for i in range(0,3):
  avgFirst+=lv[i]
letalityVaccin = []
letalityVaccin.append(avgFirst)
for i in range(3,7):
  letalityVaccin.append(lv[i])

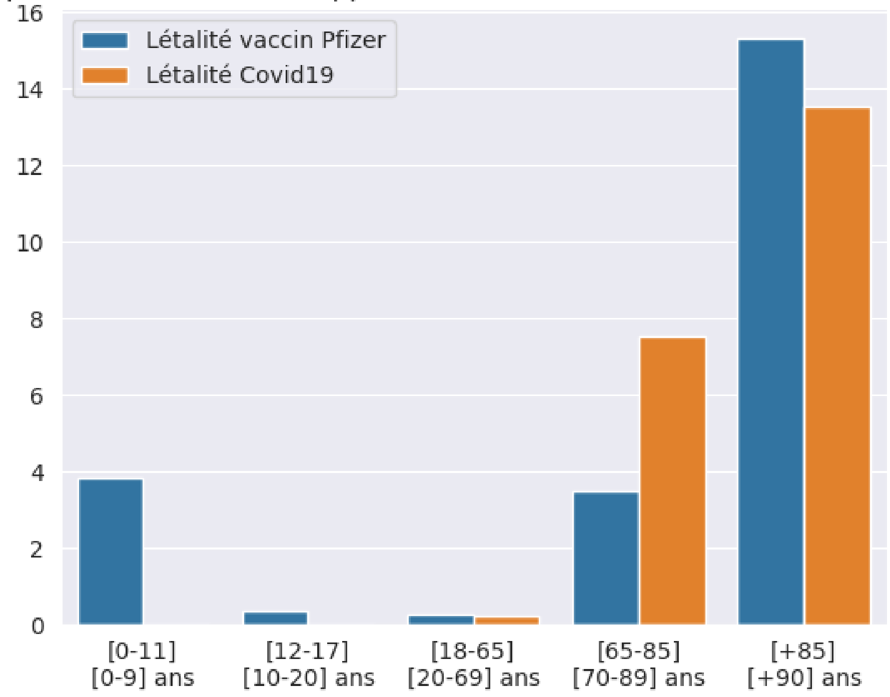
letalityCovid = []
for i in range(0,5): letalityCovid.append(resultCovidAge["Rate (%)"][i])
compare["Groupe d'âge"] = ages
compare["Vaccin"] = letalityVaccin
compare["Covid19"] = letalityCovid
ratio = []
recap.update({vaccin:letalityVaccin})
for i in range(0,len(letalityCovid)):
  ratio.append(letalityVaccin[i]/letalityCovid[i])
compare["Ratio"]=ratio
data_table.DataTable(compare)
```

[166]: <google.colab.data_table.DataTable object>

[168]:
```
raw_data_age = {
    # cat: Vaccin           Covid19
    'x': ['[0-11]\n[0-9] ans', '[0-11]\n[0-9] ans',
          '[12-17]\n[10-20] ans', '[12-17]\n[10-20] ans',
          '[18-65]\n[20-69] ans', '[18-65]\n[20-69] ans',
          '[65-85]\n[70-89] ans', '[65-85]\n[70-89] ans',
          '[+85]\n[+90] ans', '[+85]\n[+90] ans'],
    'y': [letalityVaccin[0], resultCovidAge["Rate (%)"][0],
          letalityVaccin[1], resultCovidAge["Rate (%)"][1],
          letalityVaccin[2], resultCovidAge["Rate (%)"][2],
          letalityVaccin[3], resultCovidAge["Rate (%)"][3],
          letalityVaccin[4], resultCovidAge["Rate (%)"][4],
          ],

    'category': ['Létalité vaccin '+vaccin, 'Létalité Covid19',
                 'Létalité vaccin '+vaccin, 'Létalité Covid19',
```

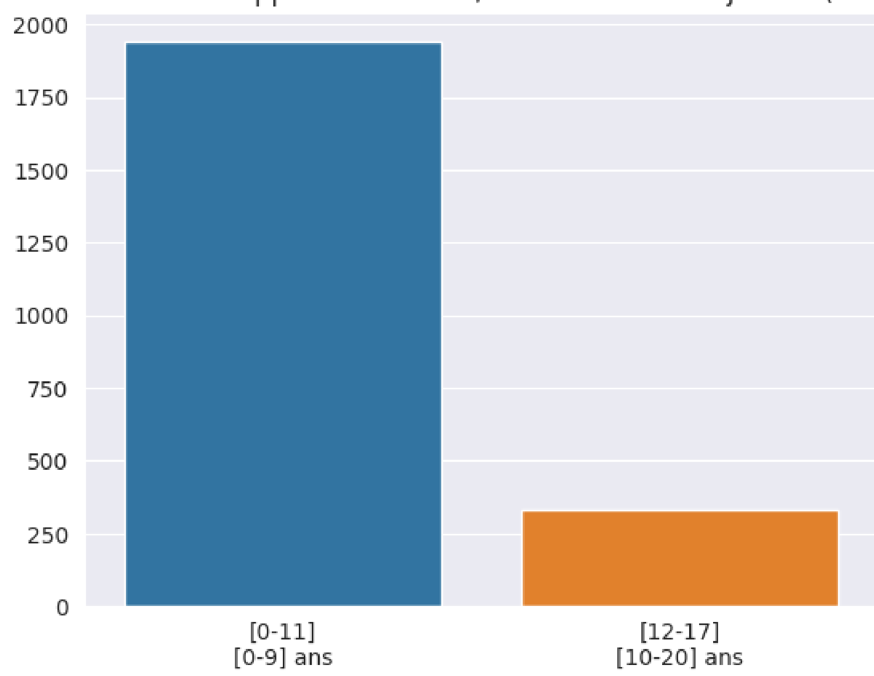
```
                  'Létalité vaccin '+vaccin, 'Létalité Covid19',
                  'Létalité vaccin '+vaccin, 'Létalité Covid19',
                  'Létalité vaccin '+vaccin, 'Létalité Covid19']
            }
sns.set_style('darkgrid')
sns.barplot(x='x', y='y', hue='category', data=raw_data_age)
plt.title("Comparaison de la létalité apparente (%) entre le vaccin "+vaccin+"␣
 ↪et le Covid19")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/
 ↪completaliteage_"+vaccin+".png")
plt.show()
```



[169]:
```
raw_data_age = {
    # cat: Vaccin             Covid19
    'x': ['[0-11]\n[0-9] ans',
          '[12-17]\n[10-20] ans'],
    'y': [letalityVaccin[0]/ resultCovidAge["Rate (%)"][0],
          letalityVaccin[1]/ resultCovidAge["Rate (%)"][1]
          ],

    'category': ['Ratio létalité vaccin '+vaccin + "/ Covid19",
                 'Ratio létalité vaccin '+vaccin + "/ Covid19",
```

```
                  ]
            }
sns.set_style('darkgrid')
sns.barplot(x='x', y='y', data=raw_data_age)
plt.title("Ratio entre la létalité apparente vaccin / Covid19 chez les jeunes␣
 ↪(vaccin "+str(vaccin)+")")
plt.savefig("/content/drive/MyDrive/MetaAnalyseVaccin/
 ↪completaliteage_jeunes"+vaccin+".png")
plt.show()
```



6 Générez et téléchargez le rapport complet

[171]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc ipyrmd
!pip install git+https://github.com/jupyter/nbconvert@master
from google.colab import output

!pip install -U -q PyDrive
from google.colab import drive
from google.colab import files
from pydrive.auth import GoogleAuth
```

```
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

# Montage du drive et création des répertoires
drive.mount('/content/drive')
!mkdir /content/drive/MyDrive/MetaAnalyseVaccin/
!mkdir /content/drive/MyDrive/MetaAnalyseVaccin/datas/
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)
downloaded = drive.CreateFile({"id":"1-m-yTOSfSZcPONa5pmmkRYEMlOwT5iYf"})
if vaccin == "Pfizer":
  downloaded.GetContentFile("/content/drive/MyDrive/MetaAnalyseVaccin/
↪Balance_Benefice_Risque_Pfizer.ipynb")
  !jupyter nbconvert "/content/drive/MyDrive/MetaAnalyseVaccin/
↪Balance_Benefice_Risque_Pfizer.ipynb" --to pdf
if vaccin == "Moderna":
  downloaded.GetContentFile("/content/drive/MyDrive/MetaAnalyseVaccin/
↪Balance_Benefice_Risque_Moderna.ipynb")
  !jupyter nbconvert "/content/drive/MyDrive/MetaAnalyseVaccin/
↪Balance_Benefice_Risque_Moderna.ipynb" --to pdf
if vaccin == "Astrazeneca":
  downloaded.GetContentFile("/content/drive/MyDrive/MetaAnalyseVaccin/
↪Astrazeneca.ipynb")
  !jupyter nbconvert "/content/drive/MyDrive/MetaAnalyseVaccin/Astrazeneca.
↪ipynb" --to pdf
output.clear()
```